Vol. XII, Part 3 August, 1960

THE
QUARTERLY JOURNAL
OF
EXPERIMENTAL PSYCHOLOGY

Editorial Board

O. L. ZANGWILL

D. RUSSELL DAVIS R. C. OLDFIELD

H. B. Barlow
D. E. Broadbent
A. Elithorn
R. L. Gregory
K. R. L. Hall
R. A. Hinde
I. M. L. Hunter
P. H. R. James
A. R. Jonckheere
H. Kay
A. M. Uttley
A. J. Watson
L. Weiskrantz
J. W. Whitfield

Editorial Assistant: Kathleen P. Watts

PUBLISHED FOR
THE EXPERIMENTAL PSYCHOLOGY SOCIETY
BY
W. HEFFER & SONS LTD. CAMBRIDGE

OCCUPATIONAL PSYCHOLOGY

Editor: ALEC RODGER

July, 1960 VOLUME 34, No. 3

CONTENTS

An experiment on the effect of specially designed safety posters. By S. Laner and R. G. Sell.

Speed and power factors of intelligence: further evidence. By O. R. Porebski.

A reply to Dr. Porebski. By D. F. Vincent.

Dr. Remitz on white-collar satisfaction. By John D. Handyside.

Food assessment and food acceptance as a psychological theme. By R. Harper.

Bias in the measurement of food preferences by triangular tests. By R. A. M. Gregson.

The development and use of a taste panel technique. By A. S. C. Ehrenberg and J. M. Chervan.

Book Reviews.

Other Books Received.

Annual Subscription 40 *shillings.*

NATIONAL INSTITUTE OF INDUSTRIAL PSYCHOLOGY
14 Welbeck Street, London, W.1.

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by BORIS SEMEONOFF

Vol. 51, Part 3. August, 1960 20s. net.

Inhibition: A Symposium.

- G. W. Pilkington and Peter McKellar. 1. Inhibition as a concept in psychology.
- R. G. A. Stretch. II: The investigation of inhibitory phenomena from the standpoint of comparative psychology.
- Donald C. Kendrick. III. Effects of drive and effort on inhibition with reinforcement.
- H. Gwynne Jones. IV. Individual differences in inhibitory potential.
- R. L. Reid. V. Inhibition—Pavlov, Hull, Eysenck.

R. Natadze. Emergence of set on the basis of imaginal situations.

J. R. Smythies. The stroboscopic patterns. III. Further experiments and discussion.

Vernon Hamilton. Imperception of phi: some further determinants.

A. H. Illiffe. A study of preferences in feminine beauty.

Publications Recently Received.

The subscription price per volume, payable in advance, is 60*s. net* (*post free*).

Subscriptions may be sent to any bookseller or to the

CAMBRIDGE UNIVERSITY PRESS
Bentley House, Euston Road, London, N.W.1.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XII 1960 Part 3

ON THE FAILURE TO ELIMINATE HYPOTHESES IN A CONCEPTUAL TASK

BY

P. C. WASON

From the Medical Research Council Industrial Psychology Research Group, University College, London

This investigation examines the extent to which intelligent young adults seek (i) confirming evidence alone (enumerative induction) or (ii) confirming and disconfirming evidence (eliminative induction), in order to draw conclusions in a simple conceptual task. The experiment is designed so that use of confirming evidence alone will almost certainly lead to erroneous conclusions because (i) the correct concept is entailed by many more obvious ones, and (ii) the universe of possible instances (numbers) is infinite.

Six out of 29 subjects reached the correct conclusion without previous incorrect ones, 13 reached one incorrect conclusion, nine reached two or more incorrect conclusions, and one reached no conclusion. The results showed that those subjects, who reached two or more incorrect conclusions, were unable, or unwilling to test their hypotheses. The implications are discussed in relation to scientific thinking.

INTRODUCTION

Inferences from confirming evidence (Bacon's "induction by simple enumeration") can obviously lead to wrong conclusions because different hypotheses may be compatible with the same data. In their crudest form such inferences are apparent in the selection of facts to justify prejudices. In research merely confirming evidence is clearly of limited value. For example, suppose that it is suggested that a deficit (x) of a particular substance in the blood is uniquely related to a distinctive symptom (y). (In logical terms, x is a "necessary-and-sufficient" condition for y.) And suppose that this hypothesis had been supported by confirming evidence alone, i.e. whenever the deficit had been induced, the symptom had appeared (ignoring statistical issues for the sake of simplification). It might then be assumed that the hypothesis was tenable. But the evidence only allows the inference that x is a sufficient condition for y, that the deficit always leads to the symptom. To establish the postulated relation, there must also be no disconfirming evidence—no case of the symptom without the deficit. For if such a case were obtained, the deficit would not be a necessary condition for the symptom, and the symptom would not be a sufficient condition for the deficit. The hypothesis that either was a necessary-and-sufficient condition for the other could be eliminated. The symptom would not be a reliable sign. Its absence would rule out the possibility of the deficit, but its presence would be ambiguous.

In general, scientific inferences are based on the principle of eliminating hypotheses, while provisionally accepting only those which remain. Methodologically, such eliminative induction implies adequate controls so that both positive and negative experimental results give information about the possible determinants of a phenomenon.

This investigation seeks to determine the extent to which intelligent young adults make rational inferences about abstract material which does not obviously suggest a conventional scientific approach. A concept attainment task would seem to be most suitable for this purpose, and one such typical task will now be considered in detail.

Bruner, Goodnow and Austin (1956) report some ingenious experiments of this kind. The material of one experiment consisted of an array of 81 instances made up from all the combinations of four attributes, each exhibiting one of three values, i.e. shape (cross, square, circle), colour (green, red, black), number of figures (one, two, three) and number of borders (one, two, three). The subject's task was to attain a concept, defined in terms of the values of these attributes, by choosing successive instances in order to find out whether they exemplified the concept. The authors describe a number of possible strategies which can be used. One of these, "successive scanning," is equivalent to induction by simple enumeration in that it involves testing successive hypotheses by using information derived from positive instances alone. For example, if "green squares" is the subject's hypothesis, instances exhibiting these features are selected. If these turn out to be positive, "green squares" is considered to be necessary-and-sufficient for the concept. But if any of these instances are negative, a different hypothesis is similarly tried out, and accepted so long as it leads to positive instances.

However, Bruner *et al.*, are not concerned with the fact that successive scanning can lead to merely sufficient, as opposed to necessary-and-sufficient, concepts. For example, if the correct concept is "red figures" (of which there are 27 positive instances), it is possible to attain the incorrect concept, "red circles" (of which there are 9 positive instances), by consistent use of confirming evidence, since all positive instances of this latter concept are also positive instances of the correct one. The incorrect concept, in this case, entails the correct one but is not entailed by it, i.e. "red circles" is a sufficient, but not a necessary-and-sufficient, condition for "red figures." Thus it is logically possible to arrive at incorrect concepts by successive scanning because only confirming evidence is utilized. Such a result, however, may not occur often because other instances, which might act as reminders of alternative possibilities, are displayed in front of the subject. In the above example, the subject's attention might be directed to the possibility of "circles" being necessary, owing to the presence of instances exhibiting red squares and crosses.

The present investigation is designed to compel the subject to encounter plausible sufficient conditions of the concept which must ultimately be eliminated in order to attain the necessary-and-sufficient conditions. The logical mechanism underlying efficient performance in the task rests upon what Von Wright (1951) terms "the fundamental, though trivial, fact that no confirming instance of a law is a verifying instance, but that any disconfirming instance is a falsifying instance."

In the experiment the concept to be attained is "*three numbers in increasing order of magnitude.*"

Subjects were told that the three numbers 2, 4, 6, conformed to a simple relational rule and that their task was to discover it by making up successive sets of three numbers, using information given after each set to the effect that the numbers conformed, or did not conform, to the rule.

It will be seen that this task, as a whole, differs from previous studies of concept attainment. In the first place, the attributes are rules referring to relations between numbers, e.g. "consecutive even numbers." In this respect, the possible concepts are more like those found in the Goldstein-Scheerer (1941) test of abstractive ability, than those which can be attained in Bruner's experiment, in which attributes cannot be combined into classes designated by names but have to be enumerated, e.g. "three red circles."

Secondly, the possible instances (triads of numbers) are, in principle, infinite. In previous studies a finite universe of instances has been used. Such a universe, however, places great constraint on the scope of inductive thinking because the number of instances which exemplify any sufficient concept will always be less than the number which exemplify any necessary-and-sufficient one. Hence the number of instances of a given kind which can be tested is limited. But in the present task an endless series of instances, exemplifying a sufficient rule can be generated without forcing the subject to encounter an instance which would not exemplify it.

Thirdly, the instances are not presented as stimuli, but have to be generated by the subject. In this way he is completely free to decide on the kind and amount of evidence which he considers adequate. If, on the other hand, all the possible instances are displayed simultaneously, the subject will know that all available evidence for the solution of the problem is already present.

Finally, no memory of previous instances is involved. The subject keeps a record of successive instances, the reasons why he generated them and their outcome. The introduction of memory into concept attainment studies is an interesting, but essentially gratuitous variable which may distort inductive reasoning by placing an additional burden on the subject.

FIGURE 1

Numbers	Reasons for choice	Conforms	Does not conform
2 4 6		√	

Record sheet

PROCEDURE

The subjects, 29 psychology undergraduates (17 men and 12 women), were examined individually and instructed as follows:

"You will be given three numbers which conform to a simple rule that I have in mind. This rule is concerned with a relation between any three numbers and not with their absolute magnitude, i.e. it is not a rule like all numbers above (or below) 50, etc.

Your aim is to discover this rule by writing down sets of three numbers, together with reasons for your choice of them. After you have written down each set, I shall tell you whether your numbers conform to the rule or not, and you can make a note of this outcome on the record sheet provided. There is no time limit but you should try to discover this rule by citing the minimum sets of numbers.

Remember that your aim is not simply to find numbers which conform to the rule, but to discover the rule itself. When you feel highly confident that you have discovered it, *and not before*, you are to write it down and tell me what it is. Have you any questions?"

Subjects then wrote down their first set of numbers, under the numbers 2, 4, 6, on the record sheet (see Fig. 1), together with the reasons why they had chosen them. The Experimenter then said "those numbers do conform to rule," or "those numbers do not conform to the rule," according to whether they were in increasing order of magnitude. The second set of numbers was then written down by the subject, the Experimenter giving the appropriate information as before. This procedure continued until the subject wrote down a rule. If the rule was the correct one, the experiment was concluded. If it was incorrect, the subject was told so and instructed to carry on as before. The experiment continued until the correct rule was announced, or the time for the session exceeded 45 minutes, or the subject expressed a wish to give up. The time was recorded and the implications of the results discussed. Finally, subjects were warned not to talk about the experiment.

RESULTS

Quantitative

Results will be classified as follows:

(*a*) frequency of correct and incorrect rules,
(*b*) extent of enumerative and eliminative thinking,
(*c*) frequency of negative instances,
(*d*) immediate response to incorrect rules, and
(*e*) types of incorrect rule.

2 4 6

(*a*) *Frequency of correct and incorrect rules.* Table I shows the frequency of successive announcements of rules. The first announcement can either be correct (defined as the "immediate correct announcement") or incorrect. (In nearly all cases incorrect rules were sufficient ones, e.g. "increasing intervals of two.") However, a third category, "none," is included for those subjects (one in this case) who made no announcement of a rule of any kind throughout the experiment. Those cases which fall into the "incorrect" category at the first announcement are redistributed within the second announcement categories, according to whether the second rule announced is correct or incorrect. But once again, the category "none" is included to cover those subjects who made no subsequent announcement of a rule of any kind after their first incorrect one. Similarly, those cases which fall into the second incorrect announcement category are redistributed within the third announcement categories on the same basis. Thus, the table provides a running record of the behaviour of the subjects.

TABLE I

FREQUENCY OF ANNOUNCEMENTS

($n = 29$)

First Announcement			*Second Announcement*			*Third Announcement*			*Fourth Announcement*			*Fifth Announcement*		
Immediate Correct	None	Incorrect	Correct	None	Incorrect	Correct	None	Incorrect	Correct	None	Incorrect	Correct	None	Incorrect
6	1	22	10	3	9	4	3	2	0	1	1	1	0	0

In the sample there were 11 undergraduates in their first year, 12 in their second year and six in their third year. Of the six subjects who made the immediate correct announcement, four were in their second year and two in their third year. There were no sex differences and no differences between Arts and Science background.

(*b*) *Extent of enumerative and eliminative thinking.* An index of the kind of thinking used was constructed by considering the nature of the instances in relation to the reasons given by the subject for choosing them. Each "reason for choice" was classified as either compatible, or incompatible, with all subsequent instances (including that instance for which each was specifically given). The total number of instances, compatible and incompatible with reasons, was computed for each subject, (i) up to the first rule announced, and independently (ii) from the first rule to the second one (if any). It was assumed that thinking was enumerative if there were a high proportion of compatible instances, and eliminative if there were a high proportion of incompatible ones.

The ratio of the number of incompatible, to compatible instances provided an eliminative/enumerative index for each subject. ① The mean ratio was 1·79 ($n = 6$) for those who made the immediate correct announcement, and 0·24 ($n = 22$) for those who made a first incorrect announcement. As predicted on the basis of the logical considerations given in the Introduction, a significant difference was obtained between these two means ($p = 0{\cdot}0002$, one tail test). The statistic used was Whitfield's (1947) extension of Kendall's S (1948) to a dichotomous variable.

The mean number of instances generated before making the immediate correct announcement was 8·0 (range 5 to 9), and the mean number before making a first incorrect announcement was 3·68 (range 1 to 7). This difference is, of course, highly significant. Thus, subjects who arrived at merely sufficient rules did so on the basis of relatively few confirming instances, while those who attained the necessary-and-sufficient rule tended to eliminate sufficient ones. ②

For those subjects whose *second* announcement was (i) correct, and (ii) incorrect, the mean ratios were 0·50 and 0·19 respectively. The difference between them was just short of statistical significance ($p = 0{\cdot}08$, one tail test).

It should be noted that in 18 out of the 22 cases of a first incorrect rule, the reason given for the subject's first instance was subsequently announced as at least a part of their first rule.

Figure 2 shows Vincent curves (Hilgard, 1938) of the mean number of incompatible instances (per fifths of the total number of instances for each subject) for (i) the immediate correct announcement, (ii) the first incorrect announcement, and (iii) the first incorrect announcement, omitting 15 subjects without any incompatible instances.

(Vincent curves represent the performance of subjects, equated in terms of their progress towards a criterion which is not defined by a fixed number of trials. The total number of trials taken by each subject to reach this criterion is divided into equal fractions, e.g. fifths, tenths, etc., and the progress made within each successive fraction is recorded.)

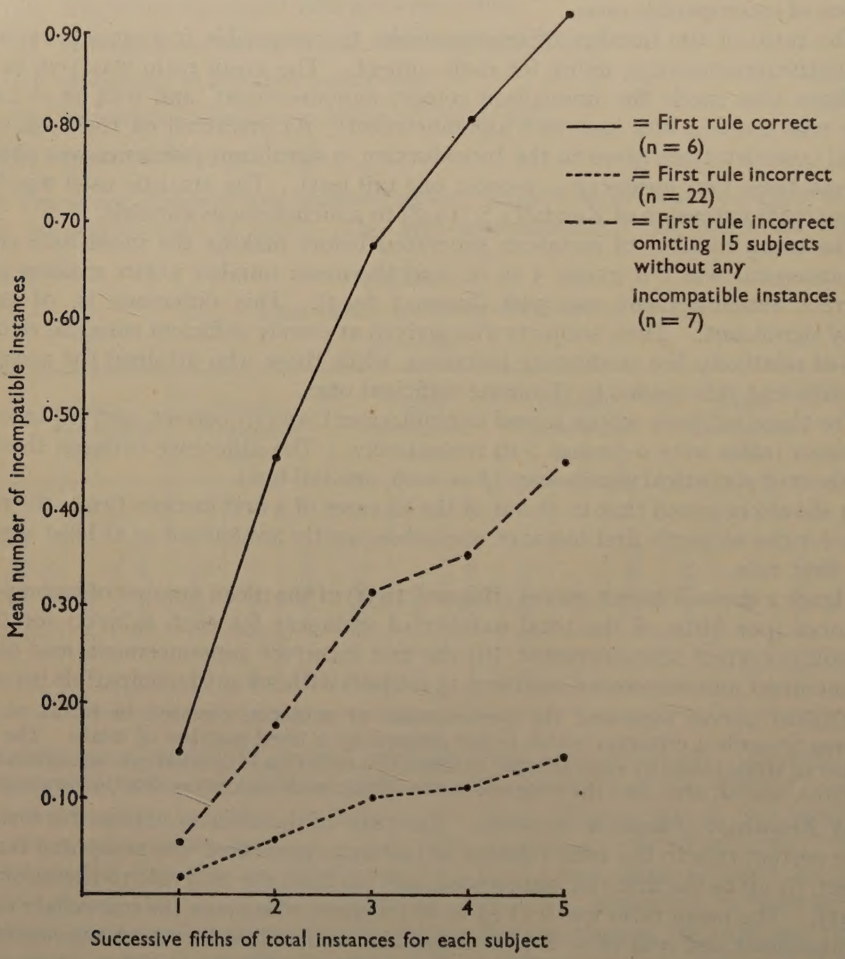
(*c*) *Frequency of negative instances.* The ratio of the number of negative instances of the correct rule to the total number of instances generated was computed for each subject, (i) up to the first rule announced, and (ii) from the first rule to the second one (if any). The mean ratio was 0·21 ($n = 6$) for those who made the immediate correct announcement and 0·04 ($n = 22$) for those who made a first incorrect announcement. As predicted, a significant difference was obtained between these two means ($p = 0{\cdot}0002$, one tail test).

For those subjects whose *second* announcement was (i) correct, and (ii) incorrect, the mean ratios were 0·38 and 0·09 respectively, and the difference between them was significant ($p = 0{\cdot}003$, one tail test).

A highly significant correlation was obtained between the eliminative/enumerative index and the negative instance index, the value of tau being $+\,0{\cdot}72$ ($p < 0{\cdot}00003$, one tail test). It should be noted that a negative instance of the correct rule does not suffice to eliminate a sufficient rule. For example, a negative instance such as 6, 4, 2 is compatible with the sufficient rule, "intervals of two in increasing magnitude." A sufficient rule can only be eliminated by a positive instance which is incompatible with it.

(*d*) *Immediate response to incorrect rules.* The instance which is generated immediately after an incorrect announcement can either be compatible or incompatible with the rule announced, and it can be either a positive or a negative instance of the correct rule. Table II shows the frequency of cases in these four categories.

FIGURE 2



Vincent curves of the mean number of instances incompatible with reasons up to the first rule.

On a purely logical criterion it would be expected that, when subjects knew that their rule was incorrect (by being told so), they would depart from it and try a new one. But it will be seen from the table that in more than half the cases the rule is maintained, even though some other attribute, e.g. order, may be tested.

(*e*) *Types of incorrect rule.* Four different kinds of incorrect rule were announced most frequently: (i) numbers increasing in intervals of 2, i.e. $a, (a + d), (a + 2d)$ where $d = 2$; (ii) increasing multiples of the first number, i.e. $a, (a + d), (a + 2d)$ where $a = d$; (iii) consecutive even numbers, i.e. $2a, (2a + d), (2a + 2d)$ where $d = 2$; and (iv) arithmetic progression, i.e. $a, (a + d), (a + 2d)$.

In addition, five rules were announced only once: (i) "the first number added to the second number gives the third"; (ii) "numbers which add up to twelve"; (iii) "ascending progression formed by adding or multiplying by a constant"; (iv) "arithmetic or geometric progression"; and (v) "the second number is the first number plus one and the third is the first number plus four." All these rules, except the last one, are consistent with the initially given instance.

1, 3, 6, 9

TABLE II

FREQUENCY OF TYPES OF INSTANCE IMMEDIATELY FOLLOWING INCORRECT RULES

	Compatible with incorrect rule	*Incompatible with incorrect rule*	*Total*
Positive instances	11	13	24
Negative instances	5	2	7
Total	16	15	31

Table III shows the frequency of incorrect rules at successive announcements. The five rules announced only once by five different subjects are classified together under "others." The frequency of correct rules is included for comparison. It will be noticed that "increasing intervals of two" is by far the most frequent at the first announcement but that there is only one case of it as a second announcement. On the other hand, there were twice as many cases of the more general rule, "arithmetic progression," at the second announcement than there were at the first. And this is obviously because the rules are not logically independent. "Consecutive even numbers" entails "increasing intervals of two," which in turn entails "arithmetic progression." Similarly, "multiples of the first number" entails "arithmetic progression" and, of course, all these rules entail the correct one, "increasing magnitude."

TABLE III

FREQUENCY OF INCORRECT RULES AT SUCCESSIVE ANNOUNCEMENTS

Rules	*Successive Announcements*				
	1st	*2nd*	*3rd*	*4th*	*5th*
Increasing intervals of two	8	1	0	0	0
Multiples of first number	4	1	0	0	0
Consecutive even numbers	3	0	0	0	0
Arithmetic progression	3	6	2	1	0
Others	4	1	0	0	0
Correct rule	6	10	4	0	1

Qualitative

Six protocols are given below, the first three from subjects who made the immediate correct announcement, and the second three from those who made incorrect announcements.

The most interesting qualitative feature of the results is that the successive announcements of three subjects consisted in the repetition of a single rule in different terms from those used on their previous announcements. In addition, three subjects reformulated their rules immediately after being told that they were incorrect, i.e. before generating any further instances. (This latter type of repetition was not counted as a new announcement in the quantitative results.)

In protocol No. 4 it will be seen that the second rule reads, "the middle number is the arithmetic mean of the other two," and the third reads, "the difference between two numbers next to each other is the same." These rules do not mean exactly the same thing. The first can be fulfilled by instances such as 5 7 9 and 9 7 5, the second by 5 7 9, 9 7 5 and 5 7 5. But, in any case, their connotations are wider than the instances given for them warrant. But the fourth rule, "adding a number, always the same one to form the next number," which follows the negative instance, 12 8 4, is restricted in scope so that it could not be fulfilled by some instances which could fulfil the two previous rules. Its connotation is now exactly that of arithmetic progression. It is not clear whether this subject, (i) appreciated the fine difference between the second and third rules, in spite of the fact that their instances conformed to arithmetic progression, or (ii) thought that there were no differences between the rules, but assumed that their expression was relevant, or (iii) thought that the rules were completely different, failing to realize that arithmetic progression was their common factor.

The following is a short extract from protocol No. 5.

> The rule is that the central figure is the mean of the two external ones.
>
> 6 10 14: the difference between the first two numbers, added to the second number gives the third; 7 11 15: to test this theory; 2 25 48: to test this theory.
>
> The rule is that the difference between the first two figures added to the second figure gives the third.

Thus, in this case it appears from the words, "to test this theory," as if the subject was announcing what she took to be a new rule, rather than merely reformulating the previous one.

The third example is found with the rule of successive multiples.

> The rule is to start with a basic number, then double it and thirdly to multiply it by three.
>
> 14 28 42; 7 14 21: the first number being a half of the second and a third of the third; 8 16 24: same reason; 9 18 27; 50 100 150.
>
> The rule is that the second number is double the first and two-thirds of the third.

It seems likely that in these cases the subjects cannot change their concepts (as judged by their instances) but change their description of them. Changing the description of a rule makes them think they have changed the rule itself: a verbal adjustment is made to satisfy the demand for a different hypothesis. These observations seem to be related to the findings of Hull (1920) and Smoke (1932) that a concept can be accurately employed before it can be correctly verbalized. Both results testify to the fact that insight into the mental processes involved in thinking is often defective. The present phenomenon is also reminiscent of magical thinking. The failure of a spell or curse can always be ascribed to some inexactitude in its utterance, rather than to its intrinsic deficiency: it would have worked, if only it had been pronounced correctly. The immediate repetition of the second rule in protocol No. 6 is particularly striking in this context.

The rule is that the three numbers must be in an ascending series and separated by regular step intervals.

The rule is that the first number can be arbitrarily chosen; the second number must be greater than the first and can be arbitrarily chosen; the third number is larger than the second by the same amount as the second is larger than the first.

EXAMPLES OF THE PROTOCOLS

The words which follow an instance are its "reason for choice." Negative instances are printed in italics.

A. Immediate correct announcement

No. 1. *Female, aged 25, 3rd year undergraduate*

12 24 36: unit figures are even and increase in two's; 8 10 12: even numbers increasing in two's; 2 6 10: even numbers increasing in four's; *6 4 2:* even numbers decreasing in two's; 2 6 8: even numbers ascending; 8 54 98: even numbers ascending; 1 17 23: ascending numbers; 1 18 23: ascending numbers; 1 2 3: ascending numbers.

The rule is ascending numbers (9 minutes).

No. 2. *Female, aged 21, 2nd year undergraduate*

3 6 9: three goes into the second figure twice and into the third figure three times; 2 4 8: perhaps the figures have to have an L.C.D.; 2 4 10: same reason; 2 5 10: the second number does not have to be divided by the first one; *10 6 4:* the highest number must go last; 4 6 10: the first number must be the lowest; 2 3 5: it is only the order that counts; 4 5 6: same reason; 1 7 13: same reason.

The rule is that the figures must be in numerical order (16 minutes).

No. 3. *Male, aged 25, 2nd year undergraduate*

8 10 12: continuous series of even numbers; 14 16 18: continuous series of even numbers; 20 22 24: continuous series of even numbers; 3 5 7: continuous series of odd numbers; 1 2 3: continuous series but with smaller intervals; *3 2 1:* reverse; 2 4 8: doubling series; *2 2 4:* two numbers the same; *6 4 2:* reverse of original numbers; 1 9 112: simple ascending numbers.

The rule is any ascending series of different numbers. (10 minutes).

B. *Incorrect announcements*

No. 4. *Female, aged 19, 1st year undergraduate*

8 10 12: two added each time; 14 16 18: even numbers in order of magnitude; 20 22 24: same reason; 1 3 5: two added to preceding number.

The rule is that by starting with any number two is added each time to form the next number.

2 6 10: middle number is the arithmetic mean of the other two; 1 50 99: same reason.

The rule is that the middle number is the arithmetic mean of the other two.

3 10 17: same number, seven, added each time; 0 3 6: three added each time.

The rule is that the difference between two numbers next to each other is the same.

12 8 4: the same number is subtracted each time to form the next number.

The rule is adding a number, always the same one to form the next number.

1 4 9: any three numbers in order of magnitude.

The rule is any three numbers in order of magnitude. (17 minutes.)

No. 5. *Female, aged 19, 1st year undergraduate*

1 3 5: add two to each number to give the following one; 16 18 20; to test the theory that it is simply a progression of two. These are chosen so that they are more complex and not merely simple numbers; 99 101 103: to test the progression of two theory, using odd numbers.

As these numbers can hardly have any other connection, unless it is very remote, the rule is a progression of adding two, in other words either all even or all odd numbers.

1 5 9: the average of the two numbers on the outside is the number between them.

The rule is that the central figure is the mean of the two external ones.

6 10 14: the difference between the first two numbers, added to the second number gives the third; 7 11 15; to test this theory; 2 25 48: to test this theory.

The rule is that the difference between the first two figures added to the second figure gives the third.

7 9 11, 11 12 13, *12 9 8, 77 75 71*.

Subject gives up. (45 minutes.)

No. 6. *Male, aged 23, 2nd year undergraduate*

8 10 12: step interval of two; 7 9 11: with numbers not divisible by two; 1 3 5: to see if rule may apply to numbers starting at two and upwards; *3 5 1:* the numbers do not necessarily have to be in ascending or descending order; *5 3 1:* could be in descending order.

The rule is that the three numbers must be in ascending order separated by intervals of two.

11 13 15: must have one number below ten in the series; 1 6 11: ascending series with regular step interval.

The rule is that the three numbers must be in an ascending series and separated by regular step intervals.

The rule is that the first number can be arbitrarily chosen; the second number must be greater than the first and can be arbitrarily chosen; the third number is larger than the second by the same amount as the second is larger than the first.

1 3 13: any three numbers in ascending order.

The rule is that the three numbers need have no relationship with each other, except that the second is larger than the first, and the third larger than the second. (38 minutes.)

Discussion

Only six of the 29 subjects gave the correct rule at their first announcement. These subjects tended both to eliminate more possibilities, and to generate more negative instances than did those who announced a first incorrect rule. Significant differences were obtained between these two groups on both criteria at the 0·0002 level of confidence.

On the other hand, the 13 subjects, who announced only one incorrect rule, presumably did so on the basis of simple enumeration, i.e. they assumed that confirming evidence alone justified their conclusions. But there are two other possible explanations of their behaviour.

Firstly, these subjects may have announced a rule in the hope that doing so would remove them from the experimental situation. This seems unlikely, however, because they all appeared highly motivated, and they expressed considerable surprise when told that their rule was not the one which the experimenter had in mind.

Secondly, they may have assumed that there could be only one rule to which the initially given instance could conform. Familiarity with the "number series" type of problem in which there is supposed to be only one correct continuation, might have induced a set for the one "right" answer. It could also be argued that the correct rule (increasing magnitude) was so trivial that students would have been reluctant to entertain it. However, the point is not that most subjects failed to give the correct rule at their first announcement, but that they adopted a strategy which tended to preclude its attainment.

But after they had announced one incorrect rule, 10 of these 13 subjects (three made no further announcement) gave the correct rule at their next announcement. And the results suggest that they did this by eliminating more possibilities, and generating more negative instances than those subjects who announced a second incorrect rule. The difference between these two groups was just short of significance (p = 0·08) on the eliminative/enumerative index, and was significant at the 0·003 level on the negative instance index. Thus, it is possible that, at the beginning of the experiment, the reinforcement of these subjects' rules by their confirming instances blocked the notion that there might be any alternative. But after this set had been broken by an incorrect announcement, they were able to eliminate any remaining alternative which occurred to them.

These possibilities can hardly apply to those nine subjects who announced two or more incorrect rules. For after their first announcement they had evidence to show that stating rules would not remove them from the situation, and that there could be more than one possible rule. Hence, it appears as if from then on they were reasoning by simple enumeration, and announcing sufficient rules, either from an inability to do otherwise, or from a preference for what Bruner has called a "direct test." In other words, they might not have known how to attempt to falsify a rule by themselves; or they might have known how to do it, but still found it simpler, more certain or more reassuring to get a straight answer from the experimenter about the correctness of their rules. This second possibility, however, seems rather remote because the method of elimination is not difficult to apply. The attempted falsification of a rule in no way depends on discovering a suitable alternative to substitute for it. All that the subject had to do is to generate an instance which is similar to previous positive ones, but does not conform to the tested rule. If the outcome is positive, the rule can be decisively eliminated. Thus, the subject has to reason that a rule will be false, if it does not cover a positive instance of the correct rule. He must, at some stage, relinquish a rule which may have been confirmed, and adopt that eliminative strategy which Bruner calls "conservative focussing" and which Mill called the method of difference.

The announcement of a sufficient rule is, in fact, the frequent result of enumerative thinking. In the present investigation, in contrast to Bruner's experiment, the use of confirming evidence alone will compel the announcement of a rule, as the only way of finding out whether or not it is the correct one. Here there are no ready-made instances displayed which might be used to correct a sufficient rule. On the contrary, instances exemplifying such a rule can never be exhausted. Thus, the experiment demonstrates the dangers of induction by simple enumeration as a means of discovering truth. In real life there is no authority to pronounce judgement on inferences: the inferences can only be checked against the evidence.

The results show that very few intelligent young adults spontaneously test their beliefs in a situation which does not appear to be of a "scientific" nature. The task simulates a miniature scientific problem, in which the variables are unknown, and in which evidence has to be systematically adduced to refute or support hypotheses. Generating an instance corresponds to doing an experiment, knowledge that the instance conforms, or does not conform, corresponds to its result, and an incorrect announcement corresponds to an inference from uncontrolled data. The kind of attitude which this task demands is that implicit in the formal analysis of scientific procedure proposed by Popper (1959). It consists in a willingness to attempt to falsify hypotheses, and thus to test those intuitive ideas which so often carry the feeling of certitude. The methodological analogue of this attitude consists in the use of increasingly stringent controls. Obviously scientific method can be taught and cultivated. But the readiness (as opposed to the capacity) to think and argue rationally in an unsystematized area of knowledge is presumably related to other factors besides intelligence, in so far as it implies a disposition to refute, rather than vindicate assertions, and to tolerate the disenchantment of negative instances. And certainly these qualities are no less important for thinking in general than the more obvious cognitive functions associated with purely deductive reasoning.

I am primarily indebted to Dr. A. R. Jonckheere. Frequent arguments with him about the logical issues involved in this research have greatly helped to clarify my ideas, and his constructive criticism of the manuscript has been invaluable. I should also like to thank Professor G. C. Drew, Dr. R. P. Kelvin and Mr. R. D. Shepherd for valuable comments. Finally, I am indebted to my subjects for the great interest and enthusiasm which they expressed in the experiment.

References

Bruner, J. S., Goodnow, J. J., and Austin, G. A. (1956) *A Study of Thinking*. London.

Goldstein, K., and Scheerer, M. (1941). Abstract and concrete behaviour; an experimental study with special tests. *Psychol. Monogr.*, **53**, No. 239.

Hilgard, E. R. (1938). A summary and evaluation of alternative procedures for the construction of Vincent curves. *Psychol. Bull.*, **35**, 282–97.

Hull, C. L. (1920). Quantitative aspects of the evolution of concepts. *Psychol. Monogr.*, **28**, No. 123.

Kendall, M. G. (1948). *Rank Correlation Methods*. London.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. London.

Smoke, K. L. (1932). An objective study of concept formation. *Psychol. Monogr.*, **42**, No. 191.

Von Wright, G. H. (1951). *A Treatise on Probability and Induction*. London.

Whitfield, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika*, **34**, 292.

EXPERIMENTAL DISORIENTATION AND CONCEPTUAL CONFUSION

BY

F. H. GEORGE and S. B. McINTOSH

From the Department of Psychology, University of Bristol

Thirty-two subjects were examined on a visual matching task. They were tested for their ability to maintain an orientation with respect to a particular direction in the horizontal plane, while being kept in circumstances designed to minimize their input of information and create thereby some conceptual confusion.

The results suggest that subjects tend to make corrections as if they were in the same position in space throughout, even though they have no necessary reason for supposing this to be true and some reasons for supposing the opposite. It seemed that non-verbal information had to be presented to the subjects in order to suppress this tendency.

Voluntary rotation of the subject from one setting to the next produced no more than chance errors, while arbitrary rotations only produced errors when cues were inconsistent, or possibly where no cues were available at all. In the cases where no cues—or minimal cues—were available, assumptions were made by the subjects about the nature of the environment. This caused both errors and correct responses for reasons that were not justified by the evidence in the hands of the subject.

Introspective reports revealed some interesting results as to the cues utilized and concepts consciously formulated by the subjects for their use.

INTRODUCTION

Since this experiment was an attempt, in part, to build upon some previous experiments by M. F. Piercy (1957) we shall briefly summarize his experiments and their results.

Piercy carried out an experiment in which subjects were confronted by a luminous arrow, pointing in one of the following directions: north, south, east, west, north-east, north-west, south-east and south-west. The subjects were then asked to perform a voluntary rotation through 180° in the darkened room and were asked to place another luminous arrow that they were then confronted by, in the *same direction in space* as the first arrow. This experiment was also carried out with some variations, such as in the use of "landmarks," but the principal results were that the subjects suffered from some considerable degree of confusion. Piercy also found that there was more confusion with respect to the fore and aft orientation of the arrow than with respect to the left-right orientation and this was thought by him to be attributable to the fact that there were two dimensions for error in the fore and aft direction and only one dimension in the left-right direction.

During an attempt to reproduce Piercy's results, which he described as "a preliminary exploration of factors relevant to the constancy of the perception of direction in small scale spatial orientation," it was found impossible to obtain any errors with normal subjects under the original conditions used by him. This was the case even when preliminary practice in a lighted room—part of Piercy's procedure—was omitted. However, there were certain differences between the procedure used here and that used by Piercy, and these we shall discuss presently.

As only five of the 28 subjects used by Piercy were "free from neurological and psychiatric disorder," while 16 were neurotic patients and a further seven were suffering from cerebral lesions, it seems reasonable to suppose that his results are only possible when the group of subjects used are retarded in the ability to form concepts. It is true that Piercy's five "normal" subjects showed no statistical differences from the "abnormal" ones, but even so they may have been of relatively low intellectual standards.

By the same token the subjects used in our experiments were all undergraduates and were generally of superior intelligence, at any rate superior to the "dull normal or above" group used by Piercy. The ease with which they performed the Piercy type of task was disconcerting, but this led us to consider the degree of confusion which it was necessary to introduce before the tasks showed systematic errors and thus to try and isolate the important factors in spatial orientation.

Now we must discuss briefly the way in which our own experiment differed from Piercy's.

As far as the results given in this paper are concerned, there was no subject who was treated in exactly the same way as any of the subjects used by Piercy. However the first condition was very similar. This first condition involved a voluntary rotation of 180° from a position inspecting one arrow to a new position inspecting another arrow. In fact, Piercy's conditions were at first used, and with some half a dozen subjects no faults were found to occur. In these preliminary experiments (many others were tried informally) the initial position of the variable arrow was kept constant, although subsequently, in the results given here, this constancy was not maintained.

Apart from these two differences, there were no other major differences between our procedure and Piercy's for the first condition, except that the blackout conditions in our case were apparently more complete. Nevertheless there will be a doubt as to how much a direct comparison of the two sets of results are possible, and we should like to lay our emphasis on these experiments as carrying Piercy's work one step further forward rather than conflicting with it. In other words a direct comparison with Piercy's results is not the major consideration in these experiments. What is quite central to our purpose is to try and discover the method by which *concepts* are formed.

It should be said straight away that our experiments were performed in the dark; the luminous arrow was all the subject was able to see and this was painted on a disc that was laid flat on a table in front of him. At the last stage of each trial the subject has to move the disc with his hands and place it in the new position as directed.

Experimental Procedure

Four conditions were used, the apparatus for each consisting of two tables approximately 2 ft. 6 in. tall and 4 ft. apart. On each table was a circular disc 1 ft. in diameter with a 6 in. luminous arrow on it.

Figures 1 and 2 show the eight arrow positions that were used. Figure 1 shows the initial orientations of the arrow relative to the observer and Figure 2 shows the equivalent orientations which were presented to him and from which he had to return the arrow to its initial orientation. Where two arrows were used, Figure 1 represents the fixed arrow and Figure 2 the variable arrow. The two tables carrying the discs were placed in a blacked out room and parallel to the walls. Approximately 3 seconds were allowed for the subject to inspect the luminous arrow, although this was difficult to control exactly owing to the room being blacked out. A swivel chair was used for conditions 2, 3 and 4. The four conditions were:—

Condition 1

A repetition of condition 1 of the experiment carried out by Piercy, with the exceptions already noted above, i.e. the subject viewed the standard arrow and then made a voluntary movement through 180° to adjust the variable arrow. No previous practice in the light was given.

Condition 2

The same arrow was used both as the standard and variable, so that the subject was facing the same direction all the time but was rotated three or four times between the presentation of the standard and variable arrow. The subject had no means of knowing that he was to finish his rotation facing the same direction as he started, indeed he was told that he would then see "another arrow," implying that there were two distinct and separate arrows in the experiment.

Condition 3

This condition was identical with the conditions of 2 except that the subject was further confused by the removal of at least one other possible cue. The experimenter *stressed* more strongly than in condition 2 the fact that the variable arrow might not be in the same position in space, and, more important, the experimenter himself changed his position during rotation so that his position could not be used as a cue (a fixed point in the environment).

FIGURE 1

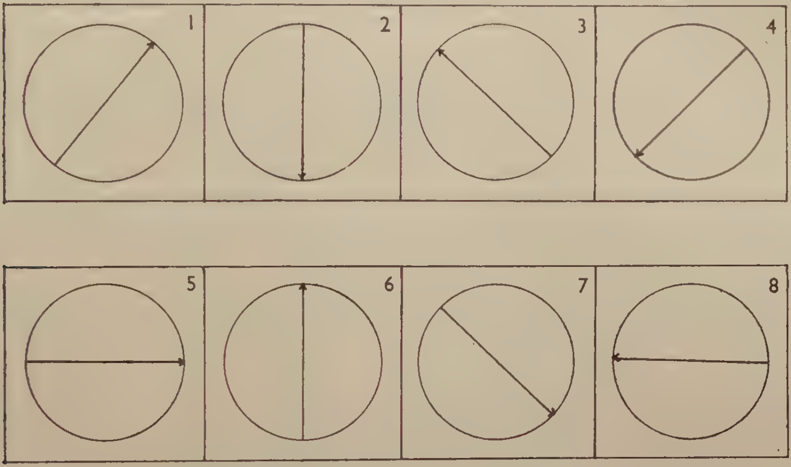
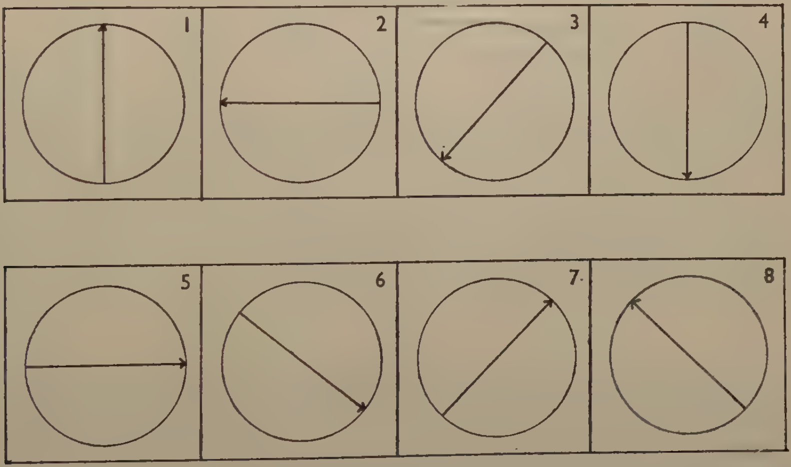


FIGURE 2



Condition 4

In half the trials two arrows were used and the other half only one in random order, so that sometimes the subject was facing in the same direction to adjust the variable arrow and sometimes in the opposite direction. The experimenter changed position as in condition 3 while rotating the subject and the emphasis on the possibility of two arrows was only as strong as in condition 2 since there really were two arrows now in half the trials.

No knowledge of results was given. All the subjects were given preliminary instructions, before entering the dark room, as follows:—

All conditions

"I am going to take you into a dark room and I want you to keep your eyes shut unless I specifically ask you to open them. (This is because I shall have to put the light on from time to time and if you open your eyes they will become light adapted and you will find the experiment more difficult). It was explained that the arrows were on tables in front of them and they were told not to touch them until they made their final adjustment to the arrow.

I will ask you to open your eyes and you will find in front of you a two-dimensional luminous arrow. I want you to note the direction in which the arrow is pointing in space, not in relation to yourself, but in relation to space, i.e. whether it is pointing towards Birmingham or Timbuctoo." Now the four conditions differ slightly.

Condition 1:—

> "I then want you to turn round, and behind you there will be a second arrow. I want you, without looking back, to point this arrow in the same direction in space as the first arrow."

Conditions 2 and 4:—

> "I will then rotate you and show you *another* arrow which I want you to point in the same direction in space as the first arrow."

Condition 3:—

> "I will then rotate you and show you *another* arrow which will not necessarily be in the same position in the room as the first arrow."

It was made quite certain that the subject understood what he was expected to do, before he was taken into the dark room. The instructions were then repeated as the subject performed each step.

For conditions 2, 3 and 4 the procedure was as follows:—

> "Open your eyes. Note the direction in which the arrow is pointing in space—Close your eyes."

When the same arrow was being used as the variable it was then adjusted by the experimenter but the light was always put on at this juncture to avoid the turning of the light on or off being a cue.) The subject was then rotated and instructed to "adjust the other arrow."

Each subject was tested with eight intercardinal directions (Figs. 1 and 2). Responses were recorded on a record sheet and at the end of the session each subject was asked:—

(i) "On what basis did you make your judgements?"

(ii) "Did you use anything as cues?"

SUBJECTS

Thirty-two undergraduates of the University of Bristol were tested, eight in each group. No attempt was made to measure the intelligence of the subjects, but it was assumed that they were more intelligent than the subjects used by Piercy whom he described as being of "dull normal intelligence or above."

RESULTS

Piercy's method of ignoring all errors of 10° or less was followed in drawing up Table I.

Condition 1. Only two errors occurred. These were both at the beginning of a series and could be explained on the grounds that the subject had not, in fact, fully understood what was expected of him, as no preliminary practice was given.

Condition 2. Only four errors occurred. In three of these cases they were the last responses in the series and were duplications of the position of the variable arrow. Both these factors suggest there may have been boredom or fatigue at this stage resulting not in disorientation but in poor memory so that the first position of the arrow was forgotten and only the second remained as a guide.

TABLE I

Position	*Condition 1*								*Condition 2*							
1	−	+	+	+	+	+	+	−	+	+	+	+	+	+	+	+
2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
5	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
6	+	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+
7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
8	+	+	+	+	+	+	+	+	−	−	+	+	−	+	+	+
Errors	1	0	0	0	0	0	0	1	1	2	0	0	1	0	0	0

Position	*Condition 3*								*Condition 4*							
1	+	+	+	+	+	+	+	+	−	+	+	+	+	+	+	+
2	+	−	−	+	−	+	+	+	+	+	+	+	+	+	+	+
3	−	+	−	+	−	−	+	+	−	−	−	−	−	+	−	−
4	−	+	+	+	+	−	−	+	−	+	−	+	−	−	+	+
5	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
6	−	+	+	+	−	−	+	+	+	−	−	−	−	−	−	−
7	+	+	+	−	+	+	+	+	−	−	−	−	−	−	−	−
8	−	+	+	+	−	−	+	+	+	+	+	−	−	+	−	+
Errors	4	1	2	1	4	4	1	0	5	4	5	5	6	4	5	4

Each column shows the eight different performances of the same subject.
+ means a success (error less than 10°).
− means a failure (error more than 10°).

Condition 3. Only one subject successfully completed the series. Seventeen errors occurred in all, only three of these being duplications of the position of the variable arrow, none of them occurring as a final response in a series. Eight errors occurred in the cardinal directions, three in the left-right and five in the fore-aft dimensions.

Condition 4. No subject successfully completed the series. A total of 38 errors occurred, none being duplications of the position of the variable arrow. Twenty-one errors were in both dimensions.

With respect to cardinal errors in this last condition, there were 10 in the left-right and six in the fore-aft dimension. Introspective reports were of some interest. Table II shows these results in tabular form.

We now show a brief statistical analysis of these results where the statistical difference between the means of the groups under different conditions was calculated for comparison purposes. The table shows the comparison between groups under conditions 1 and 4; 1 and 3; and 3 and 4. In Table III the probabilities were computed for a fiducial level of 0·05.

TABLE II

Cues used (conceptual and otherwise)	*Condition* 1	*Condition* 2	*Condition* 3	*Condition* 4
1. The subject's own person	0	5	3	3
2. A clock face	1	1	0	3
3. A compass face	0	1	1	0
4. A count made (or attempted) of number of times rotated	0	1	0	2
5. The position of the experimenter in the room	0	0	2	3
6. Parallel lines and imagined geometrical frame-works	2	0	0	0
7. The shape of the room implied by the table which was felt at the beginning and assumed in alignment with room	4	0	0	0

This table shows a list of cues used in the experiments.

TABLE III

Comparison	*t- valued obtained*	*t- value from t- tables*
1 and 4	$t = 4{\cdot}3$	$n = 16, t = 2{\cdot}145$
1 and 3	$t = 1{\cdot}4$	$n = 16, t = 2{\cdot}145$
3 and 4	$t = 3{\cdot}9$	$n = 16, t = 2{\cdot}145$

This table shows difference between means with respect to the most important groups classified according to their experimental treatment.

DISCUSSION

The main conclusion that one can draw from these results seems to be that reduction of visual cues has to be virtually complete before disorientation begins to occur in normal subjects. The findings of Asch and Witkin (1948) which were to the effect that visual cues are used in preference to tactile, proprioceptive and vestibular cues when there is conflicting evidence, do not rule out the possibility of these other cues being adequate for orientation.

Even when auditory and other cues are also removed little disorientation occurs unless the subject's *set* is inappropriate. This apparently occurs in the Ames' experiments where the set is towards a "normal environment."

It might be worth trying to measure the strength of this expectancy by preceding conditions 2, 3 or 4 by condition 1 so that the subject's immediately preceding experience might then bias him to expect a "180° environment."

From the introspective reports, it seems clear that subjects built up a sort of conceptual framework upon which they acted, even though some of the cues they had been verbally given conflicted with it. Thus subjects who were told that the second arrow might not be in the same position in space as the first still reported having the idea of a clock, or compass, etc., and using themselves as a cue, and regarding themselves as having exactly the same relation to the second arrow as the first. When the inconsistency of this was pointed out to them afterwards, subjects either said they had forgotten this information or became slightly annoyed asserting they could not have acted rationally at all, had they not *assumed* some fixed position.

In the latter case there seems to be an analogy with experimental neurosis in animals, though as there was no knowledge of results (reward or punishment) at the time, frustration did not occur until afterwards.

During preliminary investigations and in condition 2 it was found that many factors other than visual cues were utilised in orientation. The position in which the experimenter was standing, the number of times the subject was rotated are examples of such cues that were used. These additional cues perhaps explain the comparative lack of errors.

In condition 3 an attempt was made both to eliminate these cues and produce confusion by adding the statement that the subjects would be shown "another arrow which will not necessarily be in the same position in space as the first arrow." Condition 4 was similar without quite the same emphasis on the factor of a second arrow. It might have been interesting to eliminate the cues without introducing confusion and see whether confusion would arise spontaneously.

The very significant difference in errors between conditions 1 and 4 may have been due to the total absence of guiding cues and the disruption of normal concept formation by making cues misleading, e.g. the position of the experimenter did not remain constant.

The fact that the majority of the errors in condition 4 (20/39) occurred when the second arrow was placed at 180° from the correct direction and would thus have been correct had the arrow been in the *original* position suggests a relatively consistent attitude on the part of the subject.

If the hypothesis that concept formation is a primary factor in orientation is correct, this could account for the difference between our results and Piercy's results, as our subjects should be superior at "concept formation." It might also account for the fact that we found no significant difference between errors in the fore-aft dimension and those in the left-right dimension. Piercy ascribed this to "contamination between the two systems of conceptual analysis which identify the vertical and fore-aft dimensions of visual space," and this could be overcome with superior skill at forming new concepts, although here the difference in our conditions may be all-important.

We shall try, very briefly, to interpret the results of conceptual confusion in a cybernetic manner. We should, for example, guess that items were occurring in the input system which did not normally need discriminating (were generalized) but which now needed such discrimination. This seems to be possible both in the identification of objects and in the identification of relations, and in these experiments it is the latter which are important. In general, we have in the first case two events represented by collections of property names such as *abcdefgh* or *abcdefghi* which are normally identified and which now have to be distinguished. In the second case we have a relation Rab and we are simply disguising the characteristics $r, s, \ldots, w$ which define R, i.e.:—

$$R = f(r, s, \ldots, w)$$

for some function f.

Such results as these should be capable of being integrated into a hardware model or a blueprint for a hardware model (an effective theory) although no detailed attempt will be made to do so here.

We should, though, try and give some account of the significance of our notion of concept formation, and we can most appropriately describe it in terms of our briefly mentioned cybernetic model. This is not meant to imply that other interpretations are impossible, only that this seems a convenient manner in which to interpret the present results.

By concept formation we shall mean the setting up of functions *f*, *g*, . . . in th memory store such that with recognition (Uttley, 1954, 1955; George, 1956, 1957 1959; Culbertson, 1949) a comparison takes place between the input event and th events in store. If a sequence of symbols representing an event is incomplete, as i will be in general, then the problem of recognition is to ascribe a probability to th subset being a member of a particular complete set; that complete set which we cal the event (or event name) itself. Now there will be many circumstances met by ou model (it hardly matters whether we think of this as a computer or an organism which will require that a decision be made as to whether the event is a new event o not. If not, we shall say that stimulus generalization has taken place, otherwise new function is formed and we shall call this "concept formation." The condition for the decision are the same as those for a discrimination in a computer programme The crucial variable here is probably dependent upon the strength of the probability how many characteristics of the complete set seem to be missing from the subset, ho many incorrect ones are included in the observed subset, and the "value" (motivational importance) (George, 1957) of the expected next response and subsequent stimulus state.

Now in the conceptually confusing situation that was met with in Piercy's and our experiments, the implication is that unusual discriminations were called for For example, it is not important pragmatically, in the normal course of events, t distinguish the spatial frame of reference in two presentations of an arrow in a room either because the frames are obvious, visually, or because they are to be acted upon if at all, relative to the observer rather than the spatial frame. Thus in our experiments it is the relational data of the arrow with respect to its spatial frame of reference that have been destroyed, or at least disguised. As a result there was a tendency t identify two different events (stimulus generalization) or to form new concepts, and although these were sometimes arbitrary and inadequate they were in the easie cases more effective than the method used by those subjects who tended not to be able to form new concepts at all. Thus the contamination that Piercy observed was made possible because of the lack of this ability to discriminate two different sets of relations i.e. in our terminology, to form new concepts when needed.

REFERENCES

ASCH, S. E., and WITKIN, H. A. (1948). Studies in space orientation: I and II. *J. exp. Psychol.*, **38**, 325–37; 455–77.

CULBERTSON, J. T. (1949). *Consciousness and Behavior*. Iowa.

GEORGE, F. H. (1956). Logical networks and behaviour. *Bull. Math. Biophysics*, **18**, 337–48.

GEORGE, F. H. (1957). Logical networks and probability. *Bull. Math. Biophysics*, **19**, 187–99.

GEORGE, F. H. (1959). Models and theories in social psychology. In Llewellyn Gross (Ed.) *Symposium on Sociological Theory*. New York.

PIERCY, MALCOLM (1957). Experimental disorientation in the horizontal plane. *Quart. J. exp. Psychol.*, **9**, 65–77.

UTTLEY, A. M. (1954). The classification of signals in the central nervous system. *Electroencph. clin. Neurophysiol.*, **6**, 479–94.

UTTLEY, A. M. (1955). The conditional probability of signals in the nervous system. *RRE-Memorandum No.* 1109.

REDUNDANCY AS AN EXPERIMENTAL VARIABLE

BY

ALAN STANILAND

From the Psychological Laboratory, University of Exeter

This paper is concerned with the measure of redundancy derived from Information Theory, as an "objective" specification of material used in perceptual experiments. As examples for discussion it cites especially experiments by Fitts and his associates (1956), on redundancy in visual patterns, and by Miller (1958) on recall of redundant strings of letters. It offers no suggestion or comment on the use of the redundancy concept in neurophysiological models, nor does it question the concept as applied to ordinary language. Language has a long-term background, against which particular experiments (or crossword solving or word-guessing) occur: efficient guessing or utilization of probabilities can be referred to well established habits. In reduced situations, such as laboratory experiments with artificial material, this is not the case, and it will be argued that redundancy as calculated in the above experiments does not stand for a constituent of the experimental situation which can be assessed *a priori* and treated as an experimental variable comparable with a physical measure such as length. Finally, a different method of calculation, based on an alternative view of redundancy, will be put forward and related to the results of the experiments discussed.

A brief description of redundancy may be taken from Miller (1951):

> "By favouring some sequences of verbal units more than others, we effectively reduce the amount of information that is encoded per unit. . . . Suppose the various possible messages in a language consist of sequences of three digits: '1–9–3' or '8–7–0,' etc. So long as each digit in the message is unrelated to the other two, we can send any one of 10^3, or 1000 different messages. If the symbols are not independent, however, but if some fixed relation exists among them, the number of alternatives is reduced . . . our freedom of choice is restricted and the message cannot convey as much information."

Such restrictions, or constraints, on theoretical freedom have been shown to have several interrelated effects (c.f. Quastler, 1955):

(i) The actual uncertainty of the message source is reduced compared with the theoretical uncertainty. If constraints permit only 100 messages to be sent, the source uncertainty is $\log_2 100$ (6·64 bits) instead of $\log_2 1000$ (9·97 bits).

(ii) The coding of messages becomes uneconomical, e.g. 100 messages could be sent by using sequences of only two digits, giving 10^2 messages. Thus redundancy can be expressed mathematically as a relationship between two uncertainties, one for the theoretical range of messages, the other for the permitted number within the constraints. This is commonly reduced to percentage redundancy:

$$\frac{\text{H (theoretical source)} - \text{H (permitted selection)}}{\text{H (theoretical source)}} \times 100$$

In the case above:

$$\frac{\log_2 10^3 - \log_2 10^2}{\log_2 10^3} \times 100 = \frac{9{\cdot}97 - 6{\cdot}64}{9{\cdot}97} \times 100 = 33{\cdot}3 \text{ per cent.}$$

(iii) Because some sequences are excluded, internal or transition probabilities greater than 1/10 (for 10 digits) appear between digits. If, for example, 5 is always followed by 6 or 7, when 5 appears there are only two possibilities open for the next digit: $p\ (6) = p\ (7) = 1/2$, giving an uncertainty of 1 bit. By assessment of these transition uncertainties over a representative sequence of messages we can infer characteristics of the source without direct knowledge of it.

Production of Redundant Material

In each of the experiments cited, two sets of material were produced, one of which was called "random," the other "redundant" with specified percentage redundancy assigned to it. In each, comparison was made between results obtained from the two kinds of material. We shall consider the methods of producing and assigning redundancy to the material, without at first discussing the experimental contexts.

Fitts's metric figures

In an experiment by Fitts *et al.* (1956), histogram-like patterns were used in a recognition task. They were of a type called by Fitts "metric figures," produced by filling cells in a 4 × 4 or 8 × 8 matrix. The 8 × 8 range were chiefly used, but only the 4 × 4 will be mentioned as examples, because their smaller number allows a briefer analysis. Each "bar-detail" (or upright) was composed of 1–4 filled cells, so that the range of possible patterns is $4^4 = 256$, equivalent to an 8 bits task. From this range Fitts took two selections, one random from all patterns, the other a non-random selection in which bars of 1, 2, 3, 4, appeared once and once only. The number of these is 4 × 3 × 2 × 1 = 24, equivalent to 4·58 bits. This selection was called "constrained" and supplied the redundant material, its redundancy being calculated as $\frac{8 - 4{\cdot}58}{8}\ 100 = 43{\cdot}5$ per cent.

The figure 4·58 expresses the average sum of transition uncertainties within the selected patterns, which can equally well be discussed as number groups. This is shown if we break down the constrained set into the successive uncertainties for each digit. Using A, B, C, D, for the four bar-details, any item then gives:

Selection 1 (constrained):	A	B	C	D	
Transition probability:	0·25	0·33	0·5	1·0	
Equiv. uncertainty:	2·0	1·58	1·0	0	= 4·58 bits

i.e. A can be chosen from four sizes: when one has been chosen, B is limited to the other three, C to two, and finally D is determined, there being only three degrees of freedom.

If we tabulate the uncertainties for each item (as is done for another selection in Table I) we obtain a matrix of 24 rows with each row identical, the average for each column therefore being the same as the uncertainties in the rows.

The mean value is not unique to this selection and another can be shown having the same sum:

Selection 2	A	B	C	D	
Bar height:	1	1 or 2	1–3	1–4	
Transition probability:	1·0	0·5	0·33	0·25	
Equiv. uncertainty:	0	1·0	1·58	2·0	= 4·58 bits.

The component uncertainties are the same though their order is different. Like the first selection this contains 24 patterns. Thus 4·58 as a measure applies to both selections and could make no distinction between them. It can further be shown that the same figure is obtainable as a mean value for *any* 24 patterns or groups drawn from the same source. The selection set out in Table 1 was taken from a random number table with a few alterations to provide illustrative points. It can be called the near-random selection. To employ the method of decomposition into successive choices given by Shannon (Shannon and Weaver, 1949), the selection is ordered in terms of its first digits, the uncertainty calculated at each choice, and the results summed by rows and columns and averaged. Thus:

For the first digit choice we have:

$p\,(1) = 8/24$	0·528
$p\,(2) = 1/24$	0·191
$p\,(3) = 4/24$	0·431
$p\,(4) = 11/24$	0·516
$\Sigma\text{-}p \log_2 p$	= 1·666

TABLE I

DECOMPOSITION OF SELECTION 3

No.	Groups	A	B	C	D	Group totals	"Core"
1	1123	1·67	1·5	1·0	0	4·17	112
2	1144	1·67	1·5	1·0	0	4·17	114
3	1211	1·67	1·5	1·0	0	4·17	121
4	1232	1·67	1·5	1·0	0	4·17	123
5	1421	1·67	1·5	0·81	1·58	5·56	1421
6	1422	1·67	1·5	0·81	1·58	5·56	1422
7	1423	1·67	1·5	0·81	1·58	5·56	1423
8	1433	1·67	1·5	0·81	0	3·98	143
9	2113	1·67	0	0	0	1·67	2
10	3341	1·67	0·81	0	1·58	4·06	33$\underline{4}$1
11	3342	1·67	0·81	0	1·58	4·06	33$\underline{4}$2
12	3343	1·67	0·81	0	1·58	4·06	33$\underline{4}$3
13	3421	1·67	0·81	0	0	2·48	34
14	4122	1·67	1·79	0·92	1·0	5·38	4122
15	4123	1·67	1·79	0·92	1·0	5·38	4123
16	4134	1·67	1·79	0·92	0	4·38	413
17	4224	1·67	1·79	1·0	0	4·46	422
18	4234	1·67	1·79	1·0	0	4·46	423
19	4331	1·67	1·79	0·72	2·0	6·18	4331
20	4332	1·67	1·79	0·72	2·0	6·18	4332
21	4333	1·67	1·79	0·72	2·0	6·18	4333
22	4334	1·67	1·79	0·72	2·0	6·18	4334
23	4344	1·67	1·79	0·72	0	4·18	434
24	4444	1·67	1·79	0	0	3·46	44
Totals:		40·08	34·93	15·60	19·48	110·09	
Means:		1·67	1·455	0·650	0·81	4·585 bits	

Note: The "core" leaves out terminal digits of 0 bits, and internal redundant digits are underlined.

Similarly, given 1 as first digit, the second digit probabilities are:

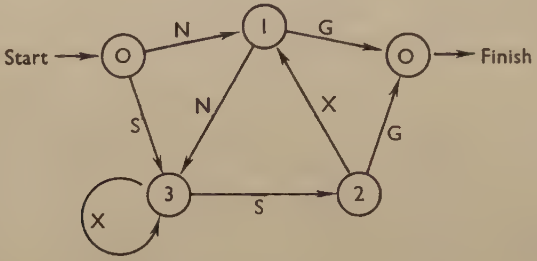
$$\begin{array}{lll} p\,(1) = 2/8 & & 0{\cdot}5 \\ p\,(2) = 2/8 & & 0{\cdot}5 \\ p\,(3) = 0 & & 0 \\ p\,(4) = 4/8 & & 0{\cdot}5 \\ \hline \Sigma\text{-}p \log_2 p & = & 1{\cdot}5 \end{array}$$

There are now three selections of the same number; each has the same average transition uncertainty; for all, calculation would give the same redundancy value relative to the source. Yet differences between them are obvious, and the question arises of how we can define the important ones. What, if anything, is indicated by their common average transition uncertainty of 4·58 bits?

Two differences stand out. In the first two selections all actual uncertainties are identical with the average uncertainties, but in the third the group sums vary between 1·67 and 6·18 bits, and all but the first column averages are different from any actual value in the column. Again, the first two selections have the uncertainty values identically ordered, whereas in the third the rank order of uncertainties varies between groups. To these differences the average figure is insensitive.

We can provisionally express this difference by saying that the first two selections are "principled," while the third is "near-random." In the first two the principle might be called "simple" in the sense that it can be briefly stated, whereas any adequate description of Selection 3 must approach very closely to an enumeration of the contents. Other selections can be made more principled than Selection 3, but whose principle cannot be as briefly stated as those of Selections 1 and 2. Thus the evidence is for a gradation of differences, rather than an abrupt distinction.

FIGURE 1



Miller's Finite-State Generator.

We can now consider the similarities. Mathematically they are not surprising. Since the only datum contributed to the redundancy formula is $\log_2 24$ for a selection of 24 groups, it is not to be expected that the result will show any discrimination between selections of that number. What Shannon's decomposition shows is that uncertainty diminishes to zero at any point where no variation is open, and is exhausted as soon as each group is uniquely defined relative to *that* selection. Thus, in Selection 3:

Given first element/s: 2, group is No. 9. Group uncertainty, 1·67 bits.
44, group is No. 24. Group uncertainty, 3·46 bits.

But, 433, group may be Nos. 19–22. The group uncertainties are 6·18 bits, of which only 4·18 have so far been resolved.

The point may be put in another way. Fitts refers to the constrained selection as "non-random," and then calculates its redundancy from the fact that it consists of 24 items. The calculation, however, applies equally to any selection of the same number of items, and the fact that his selection happens to be "principled" (i.e. non random) is an independent fact which does not enter into the measure.

Miller's redundant strings

Another form of redundant material is used by Miller (1958) in memorization and recall tasks. The source was a finite-state generator (Fig. 1), equivalent to following permitted routes on a map with uni-directional constraints (indicated by arrows). By this method Miller obtained 22 "well-formed strings" of seven or fewer letters.

> "Of these 18 were chosen in such a way that two lists of nine strings, each containing the same number of strings of the same length, could be constructed. The assignment of a string to one list or the other was determined by a table of random numbers."

Using Shannon's measure of channel capacity,* Miller obtains the value 0·552 bits/letter for the redundant strings. For random strings, based on random-numbered tables, he gives the value 2 bits/letter, from $\log_2 4$, four different letters being used in both lists.

The result of decomposition is, of course, the same for both lists, and from the facts that there are 9 strings and 54 letters can be predicted as $\frac{9 \log_2 9}{54} = 0{\cdot}528$ bits/digit. This result is not very different from Miller's for the redundant strings, which was calculated over the 22 strings produced by the generator. Discussing the well-formed strings, Miller says:

> ". . . the redundancy is 1 — (0·552/2) = 0·724, or 72·4 per cent. This can be interpreted to mean that the well-formed strings are 3·62 times as long as they would need to be to encode the same amount of information."

The random list (as will be shown later) reduces to a "core" of 20 information-carrying digits, while the redundant list reduces to only 32. This raises the question of the relationship between internal and external redundancy. Miller notes that "in this particular generator, the basic strings themselves seem to group together." And further: "The subjects seemed to distinguish only two basic types of strings." This reminds one of the readily noticeable features of the "principled" selections, and suggests that the finite-state generator may be regarded as a principle of a less simple kind than in those selections.

SAMPLES AND SOURCES

Having described in some detail two types of redundant material, we are now in a position to develop a more general argument. The aim of the psychophysical methods is, of course, to bring into the experimental situation material which can be specified by a non-psychological criterion, preferably quantitative. To use a line "3 in. long" gives complete specification, but to use a line "3 times as long" means nothing unless the comparison line is stated, and unless the ratio has a meaning within the experiment. This is also true of redundancy as a percentage measure between a sample and a source.

Fitts's statement that his constrained patterns are 43 per cent. redundant refers to all "histogram" patterns, of which there are 256. But this is not, in itself, an absolute or an ultimate figure. There is also the expansion of "all patterns made by filling 1, 2, 3, or 4 cells in each column" without constraint to the connected histogram shapes. This gives 30,625 patterns, equivalent to 15·63 bits, a source from which all the Fitts patterns could be selected, but relative to which they have different redundancy measures, the constrained selection being 70·7 per cent., the histogram selection 48·6 per cent.

If two values of redundancy are available, the choice of value requires experimental justification. This, it is suggested, must be to the effect that the chosen source has not merely been contemplated by the experimenter in the process of preparing material, but that in some definite sense it enters into the experimental situation. It will be argued that the "theoretical" values used by Fitts and Miller, both for "redundant" and also "random" material, fail to meet this requirement. Other values will be put forward, for which it will be claimed that they are more relevant values for the material as it enters into experiment.

---

* In the form $\lim_{\lambda \to \infty} \frac{\log_2 F(\lambda)}{\lambda}$ for strings of length λ or less.

We now ask, as we have asked about the redundant material, in what sense the random material bears the mark of its origin.

In the discussion of discrete noiseless channels which leads up to the definition of redundancy, Shannon (Shannon and Weaver, 1949) state that "In an ergodic process every sequence produced by the process is the same in statistical properties. Thus the letter frequencies, diagram frequencies, etc., obtained from particular sequences, will, as the lengths of the sequences increase, approach definite limits, independent of the particular sequence." Weaver similarly emphasizes that "any reasonably large sample will tend to be representative of the whole."

Clearly the sequences we have considered are not long. In fact, in Fitts's experiment only eight patterns were used in each session, a smaller selection than those here considered. We have seen that, according to its size, the random selection generates transition probabilities: it seems a reasonable suggestion that these are likely to be utilized in the successive processes of memorization and recall. Here the selection as a whole is present, and, by practical necessity, finite. This follows from the fact that it is remembered. What is suggested is that, in so far as the subject can be supposed to act in accord with a hypothesis, it is best regarded as one directed to a source identical in statistical properties with the presented sample. Or to put it more moderately, the subject's hypothesis is directed to the class of sources compatible with the presented sample. The random source will be in this class, but seems pre-eminently unsuited to be the basis of hypothesis, since it is the one source which can give no direction to response. "He adopted the hypothesis of a random source" might serve as an epitaph—perhaps as an alternative to "stimulus deprivation."

The argument started from Fitts's selections because of the clarity of the differences between them. The application to Miller's strings depends upon a re-consideration of Miller's statement that redundant material is longer than it need be to encode the same amount of information.

Miller's experiments

Miller's strings were presented on cards which were shuffled and shown at the rate of one every five seconds. After each trial the subject wrote down in any order all the strings he could recall. Ten trials were given with the same list, some subjects having the redundant lists first, others the random lists. Scoring was by correct strings, incorrect responses being ignored. The first lists for all subjects showed that redundant strings were recalled at each trial in more than twice the number of random strings. In 10 trials (on average) practically all redundant strings were recalled, but only four random strings.

The second list results varied with the type of list first learnt, and showed curious interference effects. These need not be considered as Miller's chief conclusions are based on the first trials.

In a subordinate experiment, described in a footnote, Miller's strings were used in a search task:

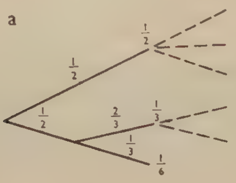
> "Lists of 30 and 60 strings were typed on separate sheets and 10 subjects were asked to search for particular strings. . . . It took about twice as long to search through the redundant strings, and it is in this sense that one can say that the redundant strings have a greater degree of similarity."

Miller comments on the similarity of this result to that obtained by Fitts with visual patterns, where significantly longer recognition times were obtained for "constrained" than for random patterns. To economize space, we shall take Miller's two experiments as representing recall and recognition situations.

Analysis by Choice Points

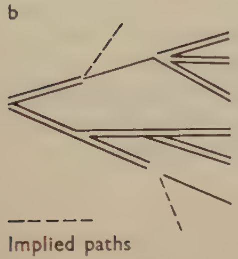
The decomposition process, as so far demonstrated, has followed Shannon's principle, but not his method of graphic representation. The diagram of Figure 2*a* precedes his derivation of the $-p \log_2 p$ formula. To it have been added only the final outgoing paths implied by his probability fractions. The decomposition replaces the probabilities with the equivalent uncertainties.

FIGURE 2.



11	1·0	1·585		2·585
12	1·0	1·585		2·585
13	1·0	1·585		2·585
211	1·0	0·918	1·0	2·918
212	1·0	0·918	1·0	2·918
22	1·0	0·918		1·918
	6·0	7·509	2·0	15·509

Average uncertainty = 2·585 bits = $\log_2 6$



111	1·0	0	1·585
112	1·0	0	1·585
113	1·0	0	1·585
211	1·0	0·918	1·0
212	1·0	0·918	1·0
223	1·0	0·918	0

The two forms are of course mathematically equivalent, giving the same analysis when applied to limited sets of items of equal probability such as appear in memorization or recognition situations. The graphic form, however, clearly displays differences of structure between sets which are less obvious in the matrix form. Each choice point (or node) corresponds to a class of similarly occurring symbols in the matrix, and the graph emphasises the number and disposition of these nodes rather than the number of separate symbols.

The subject, whether in a recall or a successive-guessing task is, in effect, tracing paths in a maze towards various terminal points. When he achieves complete recall, he can be said in some sense to know the maze. Suppose that we have six cards bearing the groups of digits of Figure 2*a* and together carrying 14 symbols. A subject who knows the set, and therefore the objective probabilities, is asked to guess at a card from the set. At the start he has to choose between digits 1 and 2, of equal probability. The first digit is uncovered as 2. He now faces an uncertainty of 0·918 bits associated with *any* digit 2 at the choice point: the fact that there are three cards starting with 2 is irrelevant, it is the class symbol 2 on whatever card it happens to appear that marks his choice point and determines his uncertainty. The uncertainty is *at* the node, compounded of the uncertainties on the outgoing paths, which following Shannon we take as the loci of probability. This suggests $\frac{6 \times 2{\cdot}585}{9} = 1{\cdot}72$ bits/path as a more relevant figure than $\frac{6 \times 2{\cdot}585}{14} = 1{\cdot}11$ bits/digit; the first figure being sensitive to differences of choice structure, whereas the second is not. One aspect of an efficient coding system is then that each symbol class (or symbol on the probability graph) represents a genuine choice point. In a modification of the graph (Fig. 2*b*), we can see the structural position of symbols which do not represent choice points. There are now two false nodes which do not correspond to choice points and are meaningless points on the effective paths. The calculated average uncertainty becomes:

$$\frac{6 \times 2{\cdot}585}{11} = 1{\cdot}41 \text{ bits/all paths}$$

$$\text{or} \quad \frac{6 \times 2{\cdot}585}{9} = 1{\cdot}72 \text{ bits/true paths}$$

and this difference is a measure of coding inefficiency in the second form. But it should be noticed that this criterion of coding efficiency can be taken still further. The *most* efficient coding system is always to have a symbol for each item, which gives 2·585 bits/true path *or* bits/symbol. Each symbol is then independent, which is (*cf.* Shannon) the mark of absence of redundancy.

With false nodes and paths we have redundancy in the sense that we are using unnecessary symbols for the task in hand: to discriminate six items. The additional nodes would be justified if there were *at least* sufficient extra items to require them. If we add two such items and make all paths true, including the two new ones, we reach the number of items which that symbol structure is fitted to deal with economically and efficiently. This would cover 8 items, and we define the *empirical redundancy* (E-redundancy) of the original structure relative to the completed structure:

$$\frac{\log_2 8 - \log_2 6}{\log_2 8} 100 = \frac{3 - 2{\cdot}585}{3} 100 = 13{\cdot}8 \text{ per cent.}$$

The structure implied by the false nodes would give:

$$\frac{8 \log_2 8}{13} = \frac{24}{13} = 1{\cdot}84 \text{ bits/paths,}$$

which is still less than the 3 bits/path we should obtain by using the symbols 1 to 8.

E-redundancy offers a measure of a characteristic of redundancy which tends to appear whenever the symbolization is unnecessarily complex for the specific choice task. It is common to all important languages, which have developed as descriptive tools for a wide range of tasks, and are always more complex than is needed for one specific task. The title "empirical" is appropriate both because it cannot be calculated from the general values, and also because the formula has an empirical aspect. (This point will be amplified later.)

A kind of redundancy arises whenever we use a code with fewer symbols than there are distinct items to be denoted. There would be economy of transcription if, for example, we had a separate symbol for each number from 1 to 99, but it would greatly increase and concentrate the discriminatory task. We reduce difficulties of discrimination at the cost of increasing symbol-events, and also spread the risk of total error. The effect is that common elements occur in symbolizations that are distinct in denotation, so that items acquire interdependence and a choice structure. E-redundancy is an attempt to describe that structure for each specific selection.

Analysis of the redundant strings

We shall now apply these principles of analysis to one of Miller's lists, L_1 "a redundant set of well-formed strings." The method of analysis by choice points is less simple in practice than the matrix method because the number of item-termini must be traced ahead, instead of appearing in the next column. At each point the value of $-p \log_2 p$ is multiplied by the number of possibilities from that point, equivalent to the addition of uncertainties in the matrix. Thus at Step 2:

$$\Sigma\text{-}p \log_2 p = 2/7 \log_2 7/2 + 5/7 \log_2 7/5$$

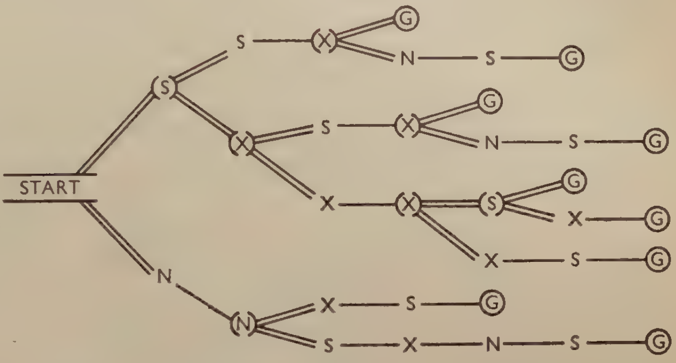
The choice point total is:

$$\begin{aligned} & 7(2/7 \log_2 7/2 + 5/7 \log_2 7/5) \\ &= 2 \log_2 7/2 + 5 \log_2 7/5 \\ &= \Sigma\text{-}n_i \log_2 \frac{n_i}{\Sigma n_i} \end{aligned}$$

FIGURE 3

STRING

S SX G
S SX NSG
S X SXG
S X SX NSG
S X XX S G
S X XX S XG
S X XX XSG
NN XSG
NN SXNSG



STEP	0	1	2	3	4	5	6	7
	9x 0·764	7x 0·864	5x 0·971 2x 1·0	2x 1·0	3x 0·918 2x 1·0	2x 1·0	0	0

Step totals 6·876 + 6·048 + 6·855 + 2·0 + 4·754 + 2·0 = 28·533
Giving bits per string = 3·17
= log₂ 9

PATHS per step:	2	3	5	6	7	6	4	= 33
No. by false nodes:	0	1	1	4	3	4	4	= 17
TRUE PATHS	2	2	4	2	4	2	0	= 16

Choice nodes are bracketed. Terminal nodes are circled. True paths are double-lined.

Decomposition of Miller's L_1 Redundant List.

The choice graph and decomposition of Miller's L_1 list is given in Figure 3. From the analysis of paths we obtain:—

True paths	16	No. of strings	9
False paths	17	Implied strings	17
	33		26
Implied paths	17	Average uncertainty of implied strings	
Total	50	$= \log_2 26 = 4{\cdot}7$ bits	

$$\text{E-redundancy} = \frac{4{\cdot}7 - 3{\cdot}17}{4{\cdot}7}$$
$$= 32{\cdot}5 \text{ per cent.}$$

Three bits/path figures can be obtained:

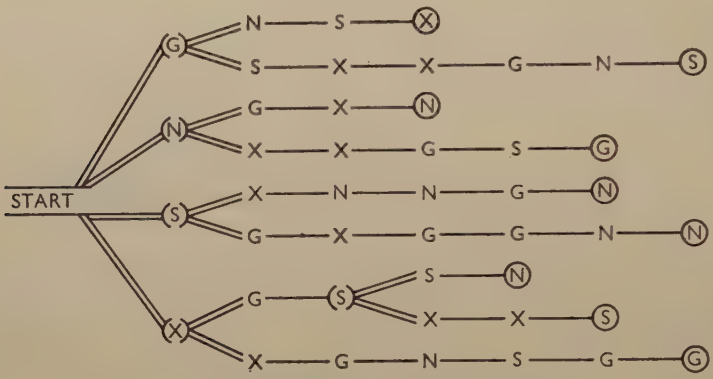
$$\text{Actual} = \frac{9\ (3{\cdot}17)}{33} \doteq 0{\cdot}86 \text{ bits/path (A)}$$

$$\text{Implied} = \frac{26\ (4{\cdot}7)}{50} = 2{\cdot}44 \text{ bits/path (I)}$$

$$\text{Reduced} = \frac{9\ (3{\cdot}17)}{16} = 1{\cdot}78 \text{ bits/path (R)}$$

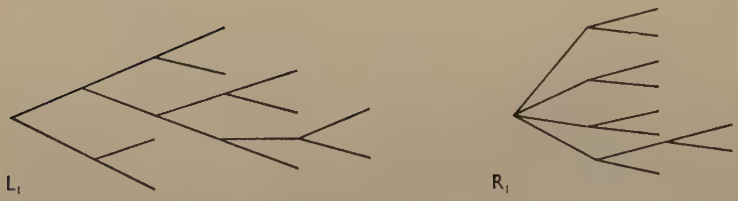
The reduced structure will give full discrimination between the nine strings without E-redundancy. The expression $\frac{\text{True paths}}{\text{All paths}}$, in this case $16/33 = 0{\cdot}484$, gives an index of reduction.

FIGURE 4



Graph of Miller's R_1 Random List.

FIGURE 5



6·87 + 6·048 + 5·61 + 6·0 + 2·0 = 17·77 + 8·76 + 2·0 = 28·53

Reduction Index: 0·484

Reduction Index: 0·311

Reduced Graphs.

The graph for R_1, one of Miller's random lists, is given without step-by-step calculation in Figure 4. Results obtained are:

$$\text{E-redundancy} = \frac{5{\cdot}322 - 3{\cdot}17}{5{\cdot}322}$$
$$= 40{\cdot}4 \text{ per cent.}$$

Bits/path figures:

Actual: 0·634 bits/path (A)
Implied: 2·8 bits/path (I)
Reduced: 2·037 bits/path (R)

By this method of measurement, the redundant L_1 list is less E-redundant than the random R_1 list. The meaning of the new measures can be indicated by comparing the reduced graphs of Figure 5. It will be seen that with exclusion of false paths the random graph reduces to three steps, but the redundant graph reduces only to five. A more efficient four-digit coding could be complete in two steps.

IMPLICATIONS OF THE RE-ANALYSIS

We may now consider the theoretical bearing of the above analysis, and we start from the characteristics of the two graphs as they appear to inspection. In the random list all discrimination points are passed in the first two steps; the remainders of the strings being entirely redundant, without choice points. In the redundant list, choice points are scattered along the strings, and, it might be suggested, provide mnemonic "stepping stones." On this basis we might put forward the qualitative hypotheses that:

(i) For memorization and recall: spreading of choice points is favourable.

(ii) For recognition or discrimination: massing of choice points is favourable.

The second hypothesis is in accord with Miller's and Fitts's results and is also intuitively plausible. In a random list, because it is random, uncertainty tends to be high in the early columns (or indeed in any column taken as starting point), therefore information for discrimination is rapidly available.

The proposed measure of E-redundancy is an empirical one in as much as there is a degree of arbitrariness in using as the reference figure for calculation the *minimum* number of items that would make all choice points genuine. It remains to offer a rational justification of this procedure, in comparison with that used by Miller and Fitts. By their procedure:

(1) A non-random (principled) selection is taken from the ensemble and assigned a redundancy based solely on the number of items obtained, without reference to the principle of selection used.

(2) A further "random" or arbitrary sub-selection is taken as experimental material and assumed to have the same redundancy as the complete non-random selection.

The aim of the present procedure is to state a redundancy for any specific selection taken from the ensemble. Now:

(1) The average uncertainty per item for a selection of n items is, as has been shown, $\log_2 n$, however the items are selected. This figure, therefore, does not give a distinctive measure.

(2) The distribution of uncertainties varies with the specific selection, so that if we could find a general description of the pattern of uncertainties this would give a measure. This we have not so far achieved. But:

(3) As the average uncertainty per item decreases the average uncertainty per digit also decreases, and when the selection is proportionately small this reduction increasingly shows itself in the form of wholly redundant digits (uncertainty $= 0$) with production of false paths.

(4) E-redundancy takes the step of treating the number of false paths as an indicant of the redundancy of the selection. The use of the logarithms of the implied and actual paths is, it may be noted, not based in Information Theory, but is merely a logarithmic scaling which appears suitable, and is open to reconsideration.

(5) Finally, in so far as the "principle" of selection shows itself in the choice graph it is by the patterning of the uncertainties. Now it can be shown (though only at some length) that the measure of E-redundancy offered varies with the factors that tend to produce spreading of the choice points, so that (as in the cases shown) high E-redundancy goes with massing of

choice points, low E-redundancy with spacing of them. It is this that suggests that E-redundancy can be used as an indicator of the form of the choice graph, though its relevance to task situations can only be established experimentally.

Experimental results

In an uncompleted series of experiments, two selections were taken from Miller's 22 redundant strings, such that their E-redundancy varied considerably. It was predicted that those with lower E-redundancy would be more easily recalled.

The conditions of the experiment were as for Miller's first trials, except that eight items were used instead of nine. This was because the same items were also to be used in a recognition apparatus similar to Fitts's, taking only eight items. Fourteen subjects were used in two groups which received the lists in opposite orders.

The average recall for the A list (E-redundancy 30·58 per cent.) was 6·92 items; for the B list (E-redundancy 37·58 per cent), 4·72 items. The individual scores were as follows:

	A list first (seven subjects)							B list first (seven subjects)						
A list:	6	8	8	7	7	6	6	7	8	5	8	7	8	6
B list:	5	7	3	3	7	3	8	4	4	4	5	6	3	3

It will be noted that only one subject gave a result contrary to the tendency of the pooled results. (A later repetition with 16 different subjects has given similar results.)

These results suggest provisionally:

(i) That a predictive distinction can be made between lists which by Miller's measure are of equal redundancy.

(ii) That the form of the choice graph has relevance.

(iii) That E-redundancy may be a useful assessment of this form.

More adequate experiments on both recall and recognition are in progress.

The opposition between Miller's measure of redundancy and E-redundancy suggests that fundamentally different concepts of redundancy are involved. These would seem to correspond to two aspects of redundancy which can be found in Shannon's formulation. In the "noiseless" channel any symbols or message units which over-determine decision are merely useless. They could be omitted or replaced with any neutral symbol. The matrix of uncertainties defines the relative usefulness (or uselessness) of symbols for the purpose of recognition, and it is this average per item which is given by $\log_2 n$.

With noise, however, symbols which are useless in the noiseless channel may acquire uncertainty-reduction values of the obscured signals which they replace. This "useful redundancy" derives from interdependence of symbols or units, and, it is suggested, from the building in of a "principle" which is not in itself defined by information measures.

E-redundancy then, makes no claim to define "principle," but only to serve as a measure of symbol interdependence or structure of choice in a finite selection of meaningless material. In the writer's view the effect of principle in memorization and recognition tasks remains to be explored by other means. The immediate conclusion is that finer specification of task material is both possible and essential. A broader conclusion, based both upon this enquiry, and upon the discussion of redundancy in visual forms by Attneave (1954), is that the word "redundancy" is carrying too many and too varied burdens, some of which touch on problems of meaning without adequately analysing them.

A last comment may be made on Miller's redundant list. As it was presented to the subjects it was meaningless and therefore only discrimination criteria can be applied to it. But it can be given meaning if we regard the strings as, e.g. prescribed racing-car routes on the finite-state map. In this interpretation no symbol is redundant for a driver, provided that he does not already know the constraints on entry to certain roads. If he knows these, or if the roads are labelled "ONE-WAY: NO ENTRY," a coding nearer to discrimination coding would suffice. Thus, finally, redundancy in a string of symbols can only be defined relative to *a priori* information and the message value for the recipient.

Error-combating redundancy is of course fully discussed by Shannon, and clearly distinguished from other redundancy. A rough criterion is that the decoded message retains language redundancies, but the transmissive redundancy is removed to obtain the message. Shannon, however, does not cover the case (common in experimental situations) of the human receiver who does not know what he is trying to receive. This remains a strictly psychological problem to which the key is probably still to be found in Bartlett's (1923) "effort after meaning."

The author is indebted to Dr. J. Szafran for much helpful discussion and criticism.

REFERENCES

ATTNEAVE, F. (1954). Some informational aspects of visual perception. *J. exp. Psychol.*, **61**, 183–94.

BARTLETT, F. C. (1923). *Remembering*. London.

FITTS, P. M., WEINSTEIN, M., RAPPAPORT, M., ANDERSON, N., and LEONARD, J. A. (1956). Stimulus correlates of visual pattern recognition: A probability approach. *J. exp. Psychol.*, **56**, 485–91.

MILLER, G. A. (1951). *Language and Communication*. New York.

MILLER, G. A. (1958). Free recall of redundant strings of letters. *J. exp. Psychol.*, **56**, 485–91.

QUASTLER, H. (Ed.) (1955). *Information Theory in Psychology*. Illinois.

SHANNON, C. E., and WEAVER, W. (1949). *The Mathematical Theory of Communication*. Illinois.

PERCEPTION OF SEQUENCE IN AUDITORY EVENTS

BY

PETER LADEFOGED and DONALD E. BROADBENT

From the Phonetics Department, University of Edinburgh, and Applied Psychology Research Unit, Cambridge

A series of tape-recorded sentences were presented to various groups of listeners, totalling 164. During each sentence an extraneous sound was present on the recording, and the listener had to indicate the exact point in the sentence at which this sound occurred. It was found that errors were made which were large compared with the duration of a single speech sound; which suggests that the listener does not deal with each sound separately but rather with a group of sounds. Errors were reduced if the sentence consisted of a series of digits rather than an ordinary text, or if the listeners were trained in phonetics. Prior knowledge of the content of the sentence did not affect accuracy. The direction of error was usually to refer the extra sound to an early point, but it is affected by the relative position of the extra sound in the sentence. These results can be regarded as an extension, to the case where all stimuli are presented to the same sense, of classic results on prior entry.

Introduction

This paper presents data which suggest that the central processes involved in understanding spoken language may have certain similarities with those involved in understanding written language. These similarities are often overlooked, because listening to speech is usually supposed to be a process in which the different speech sounds are recognized one after another. Each recognition is of course regarded as being influenced by past ones, so that a sound is more readily perceived when it is in an appropriate context. But the units on which the process is supposed to operate are single speech sounds. In reading however, it is well known that the movements of the eye do not involve fixating on each letter in turn, but instead consist of a number of rapid saccadic movements over several letters, with comparatively long pauses in between each jump. However, this method of scanning the visual data is not due to limitations of motor skill (since movements much smaller than a normal saccade can be made) nor to lack of sensitivity of the retina (which is adequately stimulated in far less than a normal fixation pause). It is a central rather than a peripheral process which is responsible for each section of the visual material being sampled for an interval of time rather similar to that necessary for the central process in a simple reaction time experiment. This being so, it would not be surprising if auditory data were sampled in a similar way. That is, if several successive sounds were handled together as a group by the mechanism responsible for identifying them. Yet if this were so the physical order of arrival of sounds might not determine their perceived order. As one sound would not be recognized and disposed of before the next arrived, all the items in each sample would be effectively simultaneous and any of them might be perceived as the first.

The experiments to be reported are not conclusive but they tend to substantiate this hypothesis. They are all based on the fact that observers find it very difficult to locate the time of occurrence of a short sound superimposed on a recording of continuous speech. The original observation made by one of the authors (P.L.), is that although he is a trained phonetician accustomed to listening to speech analytically, he was unable to note exactly when a switching click occurred in a recorded utterance which was being played back over and over again on a loop of tape.

CLICK EXPERIMENTS

Experiment I

Two groups of 12 Naval Ratings each heard a recording of a story consisting of the following eight sentences superimposed on each of which, at the point marked by a vertical line, was a click with an intensity approximately equal to the intensity of the vowel sound with the maximum stress in the sentence.

(1) There was once a young r/at named Arthur.
(2) Who could never m/ake up his mind.
(3) Whenever his friends invite/d him to go out with them.
(4) He couldn't think whether he ought t/o go or stay at home.
(5) And on the day when all the o/ther rats moved to a new house.
(6) Where they thought they'd be able to f/ind more food.
(7) Arthur couldn't bring himself t/o a decision of any kind.
(8) So he stayed behind until eve/ntually he starved.

The subjects heard a specimen click and in each sentence were told to note the words on which or between which the superimposed sound occurred. The first group did not know the story at all, and had to write down the appropriate word or words. The second group knew the story, had it written in front of them, and had each sentence read out before being played. This group noted the appropriate word or words on the text which had been provided.

Experiment II

A third group of 10 Naval Ratings heard the same recording, with instructions to mark the position of the click on a text provided. They also heard a recording of the following 10 numbers.

(1) 7 1 9 2/ 6 5 10 3 8 4
(2) 3 10 5 9 2 7 1/ 8 4 6
(3) 9/ 1 4 6 5 10 3 8 7 2
(4) 2 6 3 10 4 8 9 1 7/ 5
(5) 1 4/ 9 5 3 6 10 2 8 7
(6) 9 8 1 7 10 3/ 5 4 6 2
(7) 6 8 7 1 5/ 4 2 10 9 3
(8) 10 7 2 3 1 9 6 5 4 8/
(9) 8 2 6/ 4 9 1 7 5 3 10
(10) 4 5 8 6 3 9 2 10/ 7 1

Each number was said at an even rate with an equal stress on each digit. A click similar to those in Experiment I was superimposed on each number at the point shown by the vertical line. Subjects were asked to mark on answer sheets which were provided the digits on which or between which the superimposed sound occurred. Half the subjects heard the number recording before the story, and half received the opposite order.

General procedure and scoring

The recordings were made on Ferrograph tape recorders (frequency response 50–10,000 c.p.s. ± 2 db.) and played back on similar machines in an ordinary reverberant room. Testing was carried out in groups. Spectrograms were made to locate the superimposed items accurately. The answers of each subject were scored in the following way. Each word and each space between words was called a position. If the subject had indicated the position where the superimposed item actually occurred, his score for that sentence was zero: if he put his mark in the next earlier position, plus one was counted. If he marked the position before that, the score was plus two. Marks placed after the correct position were scored minus in a similar way. For example, in the first sentence of Experiment I, a score of plus four would mean that the word "a" had been marked, while a score of minus three would mean that the mark was between "named" and "Arthur."

Results

Experiment I. The average score per individual per sentence for the last six sentences of the story was 1·83 for Group 1 (not knowing the story in advance); and 2·3 for Group 2 (thoroughly conversant with each sentence of the story). The difference between the groups is completely insignificant statistically: but each score is highly significantly different from zero. (All subjects, except one in Group 1, had a positive score.) The click tends in fact to be heard on the average as occurring during the word before the one which it accompanies. This inaccuracy is all the more striking when we recall that each word consists of several sounds in an order which, one might think, must be perceived if the word is to be recognized. The absence of a significant difference between groups means that we have no evidence for any effect of prior knowledge of the text upon the ability to locate a superimposed item.

Experiment II. If we take the average error per sentence, ignoring sign, per subject in the third group we find that it is 2·0 for the story and only 0·7 for the numbers. This difference is highly significant, as every subject did better on the numbers. Thus there seems to be something about an ordinary sentence which makes it harder to judge the location of an item superimposed on it.

In addition, the fact that the different numbers provide examples of clicks in all possible positions, from beginning to end of the stream of sounds, allows us to check that the bias towards an early location for the click is not due to the particular locations used in the story. In the numbers, all positions of the click gave an early subjective location, with the one exception of the click in the fourth position.

We cannot therefore explain the bias towards an early location as due, for example, to a tendency to place the click nearer the middle of the range of possible positions. But in these particular subjects the earliness was more marked when the click came later: the tau correlation between the actual position of click and the size of the error in estimating the position is 0·58 ($p < 0{\cdot}05$). As this correlation is due largely to four subjects we cannot say confidently that it would appear in a fresh sample of subjects.

SPEECH SOUND EXPERIMENT

Experiment III

A possible explanation of the results given so far might be that the click used was a short and mechanical sound and not one which would in normal experience form a meaningful part of a speech sequence. Listeners might be better able to locate a sound which was exactly the same as one of the normal sounds of speech, and which could not in the first instance (i.e. before any of the sounds were identified as belonging to meaningful groups) be regarded as noise or an isolated sound having no further need of processing. The click experiments also made no attempt to study the effect of putting the extra sound in various relations to the sense group in the speech. The following experiment covers these points.

A total of 130 students heard the following set of five sentences, superimposed on each of which, in the place indicated by the vertical lines, was an [s] sound.

(Jean) He *thought* it was *Jean* w$\overset{|s|}{|h|}$o had a *cold* and was in *bed.*

(Jane) We *thought* it was *Jane* who could $\overset{|s|}{|\quad|}$ be *brave* and in the *team.*

(Jim) He *hoped* it was *Jim* who could be *k*$\overset{|s|}{|e|}$*pt* when he was *old.*

(Jack) It's *lucky* that *Jack* was in the *boat* that $\overset{|s|}{|\quad|}$ had a *leak.*

(John) They *think* it was *John* that was the *boy* that had $\overset{|s|}{|a|}$ *top.*

The prosodic features such as stress and intonation were similar in all these sentences, each being said with a heavy stress on the italicized words. The rate of delivery was that of a normal conversational style. The superimposed item was an [s] sound with a duration of 40 millisec., and comparable in intensity with the [s] in *It's* in the *Jack* sentence. (This is the only [s] in these sentences; the other sounds which are written as *s* are phonetically [z].) The recordings were made on Ferrograph tape recorders, with the aid of a Vortexion mixer unit. The overall frequency response was substantially flat over the whole speech range (± 3 db. from 50 to 9,000 c.p.s.). Sound spectrograms were made of the sentences, so as to check that each superimposed item was in exactly the intended position. Where an item was to occur on top of a word, care was taken to ensure that part of the vowel in the word occurred on either side of the [s]: and when it was to occur between words it was arranged to be during a stop closure. This closure belongs phonetically to one word or another; in ordinary speech there are seldom gaps marking words boundaries. But a stop closure such as that of the [t] in the phrase *that had* results in a period of silence after the last audible part of one word and before the first audible part of the next word: an item superimposed during such a closure may thus be said to occur between words.

TABLE I

ANALYSIS OF VARIANCE FOR THE FIVE GROUPS RECEIVING SENTENCES IN DIFFERENT ORDERS IN EXPERIMENT III

Sources of variance	*d.f.*	*Sum of squares*	*Variance*	*F*
Between Groups	4	69	17·25	<1
Within Groups	55	3429	62·34	
Total	59	3498		

TABLE II

ANALYSIS OF VARIANCE FOR THE GROUP OF TRAINED LISTENERS AND THE 60 UNTRAINED SUBJECTS IN EXPERIMENT III

Sources of variance	*d.f.*	*Sum of squares*	*Variance*	*F*
Between Groups	1	336	336	5·7
Within Groups	67	3952	58·98	
Total	68	4288		

TABLE III

THE MEAN JUDGEMENTS OF THE GROUP OF TRAINED LISTENERS AND OF THE 60 UNTRAINED SUBJECTS IN EXPERIMENT III

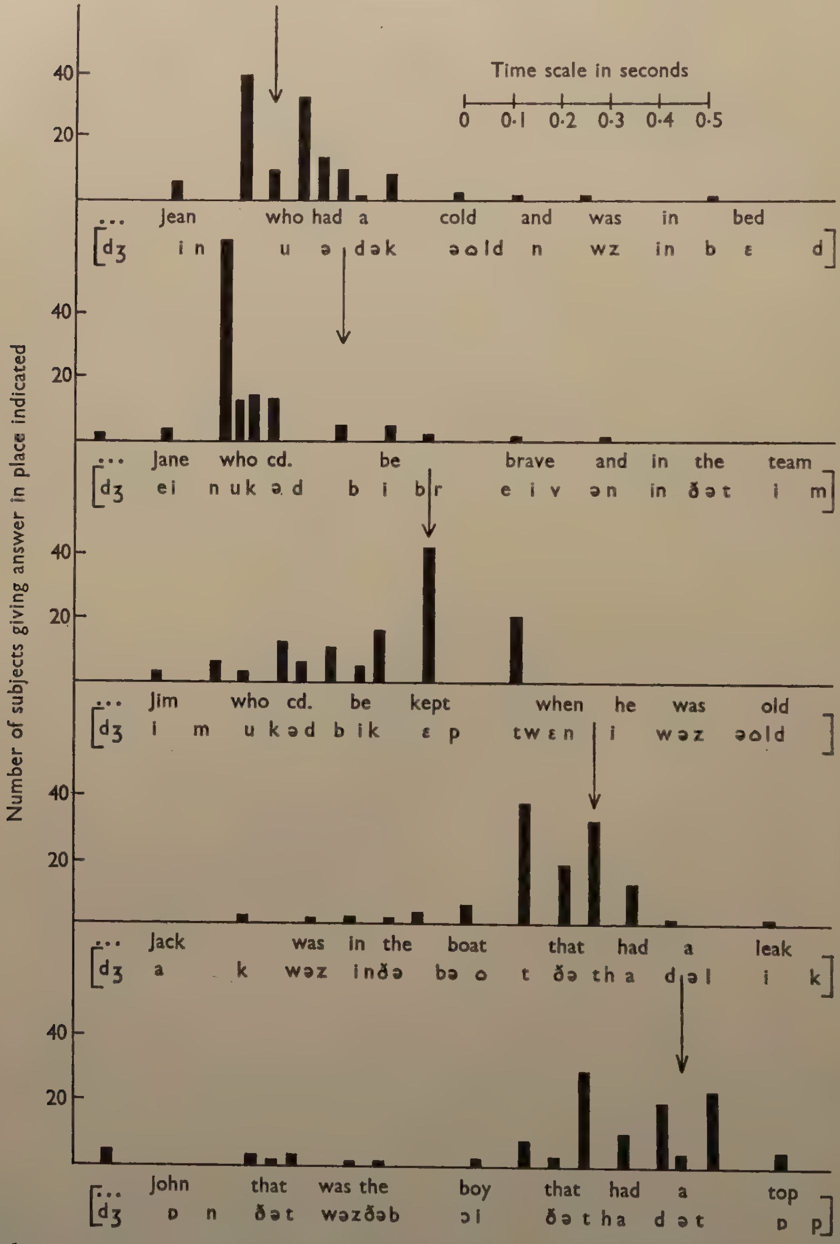
No. of subjects	*Composition of group*	*Type of sentence*					*Group totals*
		Jean	*Jane*	*Jim*	*Jack*	*John*	
9	2nd year speech therapists	−3	+2	+1	0	0	0
60	Subjects in Table I	−1	+3	+1	+1	+2	+6

All the subjects in this experiment were given a text of the five sentences which they read while first listening to the sentences without any superimposed items. They were then given written and duplicate oral instructions explaining that they were going to hear the same recording again, but this time there would be an [s] sound superimposed on each sentence; and they were to mark on the provided texts the words on which or the words between which these sounds occurred. They were instructed not to read the text at the same time as they listened to the recordings.

The 130 subjects were considered in different ways.

(*a*) 60 of them were divided into five groups, each of which heard the five sentences in a different order forming a Latin Square. (*b*) Nine of the remaining subjects were students who had been taking classes in Phonetics for over a year. (None of the other subjects in

FIGURE 1



The judgements of 121 subjects on the five sentences. The actual position of the extra sound is indicated by an arrow any of the experiments had been trained to listen to speech analytically.) (*c*) The remaining 61 subjects who were not used for statistical analysis all heard the sentences in the order *John, Jane, Jean, Jack, Jim.*

RESULTS

(*a*) *The five groups who heard the sentences in different orders.* For each subject four scores were obtained by taking away his score on the first sentence from that on the second, third, fourth, and fifth.

A multivariate test of the equality in different groups of these four scores was then performed. (This test, which is based on Hotelling's T^2, was performed by Dr. M. Stone, of the Applied Psychology Research Unit). The test showed that the four scores were different in the five groups, $p < 0{\cdot}01$.

This must mean either that the different sentences are different in the type of error they produce, or else that the different sentences produce different amounts of practice. That is, either a sentence such as *Jim* produces a different error from some other sentence, or else judging the *Jim* sentence after the *Jean* sentence is different from judging the *Jim* sentence after the *Jane* sentence. The latter possibility was tested by taking the total score for each subject, counting all five sentences. Analysis of variance was then performed on these total scores, as shown in Table I. The groups were not significantly different. Thus the total score on all five sentences is not different if they are heard in different orders. Therefore the previously proved difference in the four scores in the five groups is likely to be due to the five sentences producing different types of error.

(*b*) *The group of trained listeners.* The total score for each subject in this group was determined, counting all sentences. The difference between the group and the 60 other subjects already described was then tested as shown in Table II. This group is therefore different from the others ($p < 0{\cdot}05$). The mean judgements of this group and the other 60 subjects are shown in Table III. It seems that this group is different in that it is often slightly more accurate.

(*c*) *The remaining* 61 *subjects.* The responses of these subjects taken together with those of the 60 subjects already discussed are shown in Figure 1. The words are displayed (in orthography and in a phonetic transcription) in accord with their duration in the utterance.

DISCUSSION

All these experiments show that subjects have a great deal of difficulty in identifying the time of occurrence of some auditory events relative to others; and that they often consider the item which they are trying to pick out to come earlier than is actually the case. This situation is suggestively reminiscent of the early experiments on "prior entry" in which, e.g. subjects are told to watch a pointer moving around a drum, and to listen for the sound of a bell. In these circumstances subjects usually judge the bell to have sounded at a point which was passed by the pointer somewhat earlier than the actual point at which the bell was struck. This effect is not solely due to the shorter sensory conduction time of the ear since if subjects are instructed differently, and do not expect to hear the bell, they will, if asked afterwards, consider that it was struck at a point somewhat later in the pointer's movement than was actually the case. Titchener (1909) formulates the resulting "law of prior entry" as follows: "The stimulus for which we are predisposed requires less time than a like stimulus, for which we are unprepared, to produce its full conscious effect."

All the early work on this "law" involved stimuli affecting different sense organs, and it is of course easy to think of information in one sensory channel as being delayed relative to that in some quite different channel. But if the phenomenon of prior entry applies even to items all arriving by one sense, we must conclude that these items do not pass along the sensory paths in rigid succession and to that extent the concept of sensory sampling is justified. It is therefore worth assessing the extent to which the results described in this paper can be considered as being similar to the older experimental data on prior entry. In particular, can the shifts in location of the extra item, produced by changes in the speech material, be regarded as similar to the changes produced by "predisposition" in the classical experiments?

Some of the results can be reconciled with the earlier ones. It is a plausible hypothesis that the extent of a subject's predisposition to respond to a class of stimuli is related to the amount of attention which he is giving to that class of stimuli; and this will depend on his estimate of the probability that some member of the class is about to occur which will convey information essential for a response. Subjects will be paying more attention to a group of stimuli when there is a high probability that an information bearing stimulus is about to occur. Thus reaction time is shorter when a stimulus is presented at a time when it is highly probable (Mowrer, 1940); and conversely reaction to visual stimuli is slower the more information is conveyed by simultaneous auditory signals. Now in Experiment II we compared meaningful sentences and sequences of digits as the speech material. We are handicapped by a lack of knowledge about the amount of information conveyed by ordinary sentences such as those in the story. But there are certainly marked sequential dependencies in English sentences which are lacking in the digit series; and the sentences are easier to remember than the numbers. It is therefore probable that the sentences do not carry information at as great a rate as the series of digits, in which lexical information can be conveyed at the rate of 3·1 bits per word. Consequently, when decoding the information in the sentences in the story, the subject does not pay as much attention to the words in the utterance as he does when decoding the information in an utterance consisting of a series of digits. In the latter case his predisposition, which is normally towards the superimposed item, will be reduced: and therefore it is reasonable that he will be more accurate when the speech material consists of a series of digits.

The other finding in Experiment II was that the subjects studied showed a larger error when the click was later in the series of digits. This may plausibly be explained by the rise in the probability of the click as each series continued, since there was bound to be a click at some point, and if it had not happened in the early stages it was certain to come in the later ones. Thus the listener would be more predisposed towards the click when it came late. An alternative theory might be that this result was due to the central tendency of judgement, common in psychological situations. This would not exclude the operation of prior entry but be additional to it.

Similarly some aspects of Experiment III are in agreement with a modified prior entry theory. Each of the sentences in Experiment III was constructed so that it had three sense groups, each ending with a stressed word. In order to decode the information conveyed by such sentences a great deal of attention will be needed both at the beginning of each sense group, and at the moment of occurrence of stressed words, since at these times the message is relatively less predictable. The *Jean* sentence is the only one in which the extra item occurs on the first word in a sense group; and it is also the only one in which a considerable number of subjects considered the item to have come later than it actually did; so perhaps at this moment in the sentence subjects were paying more attention to the words than at other moments.

When the superimposed item comes on a stressed word (i.e. in the *Jim* sentence) presumably many subjects are paying about equal attention to the words and to the extra sound for which they are listening, since the *Jim* sentence was the one which many (42) subjects got right. It is interesting to note in connection with this sentence that there is a rule of English prosody: *stresses tend* to re*cur* at *reg*ular *in*tervals of *time*. Thus if the last half of the previous sentence is read aloud in a normal conversational style, while tapping a pencil on the table at the moment of saying each stressed (italicized) syllable, it will be found that, almost irrespective of the number of unstressed syllables between each of the stressed ones, the taps recur at regular intervals. Consequently, in many English sentences, after hearing the first two or three stresses the listener can approximately predict the moment of occurrence of the subsequent stresses; so the times when the maximum information will occur, and hence the times when the maximum attention will be needed, are to some extent known in advance.

It should be noted that the results of Experiments II and III argue strongly against a possible alternative to the "prior entry" explanation. This alternative view is that speech takes longer to be perceived than an isolated sound does, because it involves more complex processes. If this were so we would expect the superimposed sound to be regarded as coming even earlier when the simultaneous speech is less redundant, since the perception of the speech should be even further delayed (Hick, 1952; Hyman, 1953). But on the contrary the location of the superimposed sound appears later and so more accurate when the speech is less constrained. Thus the error is not likely to be caused by delays due to the complex character of speech.

The most serious difficulty, in relating these results to earlier prior entry effects, lies in Experiment I. The finding in that experiment was that knowledge of the text of the story in advance did not make any significant difference to the degree of difficulty of identifying the relative time of occurrence of the superimposed item. If knowledge of the text reduced the amount of information decoded when listening to the story, we would have expected one group to be more accurate than the other. This difficulty can perhaps be met by recalling that there is no reliable estimate of the amount of information conveyed by a spoken message as compared with the amount of information conveyed by the same words in written form or spoken by another voice. It is quite probable that in the conditions of these experiments prior knowledge of the lexical structure of the text has little bearing on the amount of information conveyed by the utterance as an auditory event: knowledge of the words that a speaker is going to use is only of slight assistance in predicting the actual noises—including the rate of speech, the stress pattern, the intonation and the voice quality that are going to occur. So it is perfectly possible that the subject has to attend to the sounds of the words to an equal extent, whether he knows the text or not.

Another and more minor difficulty is that the differences between the *Jane, Jack* and *John* sentence of Experiment III cannot readily be accounted for on a modified prior entry view. However, the interesting fact, which is applicable to all the sentences in all the experiments, is thoroughly established: subjects often have great difficulty in assessing the relative time of occurrence of different auditory events, even when they have been specifically asked, before hearing the events, to try to do so. This difficulty is present both when the auditory events are of different kinds (i.e. speech sounds and a click) and when the auditory events are all of the same kind (i.e. all speech sounds).

We may conclude from this that the process of decoding information when listening to speech may involve operating on units which are somewhat larger than the duration of a single speech sound. Admittedly it is not possible to say much more about the units, for there is no definite maximum error in the location of superimposed items. The auditory samples cannot therefore begin and end at the same place for all listeners to a particular set of sounds. Thus the auditory psychologist is in a similar position to that of his visual colleague. Carmichael and Dearborn (1948) says: "From the study of a great many eye-movement records the conclusion is drawn that the eye seems to have no particular preference for fixating any special part of a word or type of letter or length of word." It appears that reading and listening are similar activities in that they both normally involve shifts of attention in which there is a large random component. Despite this randomness, it is certain that reading is not a letter-by-letter process: and equally listening does not deal successively with each isolated speech sound.

References

Carmichael, L., and Dearborn, W. F. (1948). *Reading and Visual Fatigue.* London.

Hick, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.*, **4**, 11–26.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, **45**, 188–96.

Mowrer, O. H. (1940). Preparatory set—some methods of measurement. *Psychol. Rev. Monogr.* No. 52.

Titchener, E. B. (1909). *Lectures on The Experimental Psychology of the Thought-Processes.* New York.

A CONTRIBUTION TO THE STUDY OF PHENOMENAL CAUSATION*

BY

D. G. BOYLE

From the Department of Psychology, University of Hull

In his studies of the perception of causality, Michotte describes a number of "impressions," concomitant with precisely defined experimental conditions, of which "Launching"† and "Triggering"† are two of the most important. Two major factors influencing the nature of the impressions are the relative speeds of the two rectangles representing objects participating in dynamic inter-relationships, and the length of path of the more passive of the two objects. This investigation was an attempt to discover the quantitative relationships obtaining between these two independent variables and the dependent variable of the subjects' responses to the experimental conditions.

INTRODUCTION

By means of an experimental device to be described in the following section, Michotte arranges for an observer to view two rectangles, of sides 5 by 8 mm., of which one (black) appears to move across the field of view from left to right, and make contact with the other (red), which up till then has been stationary in the middle of the field. After being struck, the red object moves away in the opposite direction, either at a lower speed than the black object approached it or at a higher speed.

The colours of the rectangles are irrelevant to the impression. Whatever colours are used, the object that first approaches from the left is known as object A, and is regarded as the more active of the two. The object that moves away after being struck is known as object B, and is regarded as the more passive.

Briefly, if the speed at which A approaches B is greater than the speed at which B withdraws after impact, the impression is that A has "launched" B into motion. This impression Michotte calls "Launching." If, on the other hand, B withdraws after impact at a speed greater than that at which A approached it, the impression is that A has "triggered" B into motion. This impression Michotte calls "Triggering." The combinations of speeds, and the resulting modifications of impressions, are very great, but we have described the simplest cases as they affect the experiments to be reported.

Generally speaking, as the length of B's path increases, the impression is more likely to be Triggering than Launching, even if the ratio of speeds A : B is greater than unity (i.e. when A approaches B more quickly than B subsequently withdraws.) This case will be known as a "descending ratio" of speeds. When the length of B's path is fairly short, the impression is more likely to be Launching than Triggering, even when the ratio of speed A:B is less than unity (i.e. when B withdraws after impact more quickly than A approached it.) This case will be known as an "ascending ratio" of speeds.

These observations refer to preliminary experimentation conducted at Louvain by Professor Michotte and the present author during the session 1956–57. They are related to the concept of the "radius of action," that is, the distance that B may

* This work was conducted at Louvain, under the direction of Professor A. Michotte, during the session 1956–57, and was first reported in the author's Ph.D. thesis presented to Hull University in July, 1958.

† The terms used in this paper are those adopted by Michotte's translators, T. R. Miles and Mrs. Miles, of the University College of North Wales, Bangor.

travel while still appearing to be under the influence of A. Beyond a certain point, B's movement becomes phenomenally autonomous. This point cannot be discussed at any length here; the reader is referred to Michotte (1954) and to the experimental studies of Yela (1954) for further details. The concept must be mentioned, however, as the radius of action put an upper limit on the length of path that we were able to use in the present experiments.

Altogether, about 80 subjects took part in the present set of experiments. Of these, 40 were selected for the major experiment. They were asked to respond (in the method to be described in the following paragraph) as the length of B's path was systematically increased or decreased under the conditions of either an ascending ratio of speeds or a descending ratio (four sets of conditions altogether.)

The nature of the experimental response differed in the present major experiment from the nature of the responses in Michotte's experiments. Michotte placed no restrictions on the nature of his subjects' responses, telling them to describe what they saw "as if they were telling a friend something that they had seen happen in the street." For the present experiment, the subjects were asked to compare the experimental stimulus with two standard stimuli, one of which unequivocally gave them the impression of Launching, whilst the other unequivocally gave the impression of Triggering.

The experimental hypothesis was that, in the case of an ascending ratio of speeds, lengthening B's path would tend to reinforce the forces of organization, leading to a greater tendency to report Triggering at greater lengths of trajectory. In the case of a descending ratio of speeds, increasing the length of trajectory would lead to a conflict of the forces of organization, leading to ambiguity of response at greater trajectory lengths.

Apparatus

A full description of the apparatus is given by Michotte (1954), and summarised by Vernon (1952); a detailed description would take a considerable amount of space, and would require diagrams. Since such a description would be out of place in this paper, and since details can be obtained easily from Vernon (1952), we shall describe the apparatus only briefly.

The basic apparatus consists of a circle of thick white card, 48 cm. in diameter, on which are drawn two bands 8 mm. in width; one of these is black, the other red. (In practice, two discs are usually employed, each cut along a radius and one set of bands drawn on each. These can be fitted together, making it possible to employ different combinations of speeds without the necessity of drawing a new disc each time.)

The discs are then mounted on plates of the same diameter, revolving at constant speed in the vertical plane behind screens. In each screen a slit is cut, which is 5 mm. high and about 150 mm. long. The observer, viewing the apparatus from a distance of 1·5 metres, sees each band through the slit as a rectangle of side 5 by 8 mm.

If the bands are drawn along arcs of constant radius, the observer will see stationary "objects," even when the discs are revolving. If, however, the arcs are of decreasing radius, then, when the discs revolve, the objects will appear to approach the centre of the slit; if the arcs are of increasing radius, the objects will appear to move away from the centre. In this way it is possible to create the appearance of object A approaching object B, which is stationary at the centre of the slit, then hitting B, after which B moves away, A then remaining stationary.

A series of preliminary experiments was conducted to discover the speed ratios and lengths of path that would yield the clearest-cut results. As well as the stimulus conditions chosen on the results of these investigations, two other sets of stimuli were necessary.

The first of these has already been referred to, viz. the stimuli that gave unequivocal impressions of Launching and Triggering to the subjects taking part in the major experiment. We shall refer to these as "Prototype stimuli." The other set was necessary because most of the subjects employed were naïve to this type of experiment, and in many cases to psychological experimentation in general. It was therefore necessary to explain to them the *type of response* that was required in this series of experiments. Consequently preliminary training was necessary to ensure that subjects were attaching similar meanings to the permitted experimental responses.

The stimulus conditions were as follows.

Prototype stimuli

Speed of A (cm./sec.)	4·5	Speed of B (cm./sec.)	30·0
” ”	30·0	” ”	3·0

Training stimuli

Speed of A (cm./sec.)	18·0	Speed of B (cm./sec.)	4·5
” ”	9·0	” ”	12·0

Experimental stimuli

Speed of A (cm./sec.)	45·0	Speed of B (cm./sec.)	30·0
” ”	13·5	” ”	30·0
” ”	54·0	” ”	22·5
” ”	9·0	” ”	15·0
” ”	30·0	” ”	15·0
” ”	6·0	” ”	15·0

The apparatus was arranged with a movable flap fitted to the slit in the screen, so that the length of B's path could be increased or decreased by the experimenter. A scale was drawn on the screen so that the path could be altered by fixed steps of 5 mm. This scale was drawn so faintly as to be invisible from the subjects' viewing position, 1·5 metres in front of the apparatus, especially when this was frontally illuminated.

A number of revolving plates were employed, so that the discs did not have to be changed with each change of stimulus condition. The stimuli not being presented were covered until required by movable screens. The subject moved from one stimulus condition to the next, viewing each frontally from the same distance. All discs were clearly illuminated by frontal illumination.

Procedure

(1) *Training and pilot experiments*

All naïve subjects were presented first with the prototype stimuli and asked to describe their impressions. No specific instructions were given as to the aspects of the situations on which to concentrate. Subjects who described the prototype stimuli in the way expected from Michotte's reports (i.e. those who described the stimulus with A : B = 30·0 : 3·0 in terms appropriate to Launching, and the stimulus with A : B = 4·5 : 30·0 in terms appropriate to Triggering) were selected for training. It was found that about 50 per cent of the subjects had to be rejected on the basis of this preliminary screening test, either because their replies were completely ambiguous, or because they described the stimuli in ways opposite to that expected.

The selected subjects were then shown one of the training stimuli with object B's path reduced to 5 mm. (the path was measured from the edge of B nearer the slit, so subjects saw B as a red square of 5 mm. side, with no free space behind it.) They were asked to say which of the two prototype stimuli the presented stimulus resembled more. If they could not make up their minds they were allowed to say "uncertain," "in between" and so on.

After subjects had responded, the length of B's path was increased by the experimenter to 15 mm. and subjects were asked to make a further judgement of "more like the one on the left" or "more like the one on the right," and so on. Subjects also made judgements with the length of B's path equal to 25, 35 and 40 mm.

This procedure was repeated with the other training stimulus, after which the whole procedure was repeated with the lengths of path reduced by similar steps from 40 to 5 mm.

In these training procedures, the steps employed were a little elastic, as the idea was to acquaint the subjects with the *nature* of the task rather than to give them practice at the more difficult task of the experiment proper.

Subjects then proceeded to the pilot experiment, which was similar in principle to the experiment to be reported in the following section, although different speed values were used. The pilot experiment will not, therefore, be described, as its main function was to enable the experimenter to design the final experiment.

(1) *Main experiment*

Altogether 40 subjects took part in the present experiment. Of these, nine had taken part in the pilot experiment previously referred to. These, therefore, had received training in the task. Because we were, as a consequence, using fairly well-practised subjects, we thought it legitimate to employ also the well-trained laboratory personnel as subjects (investigation of their results showed later that these did not differ in significant respects from the results of other subjects). Besides the laboratory personnel and the subjects from the previous experiment, 26 new subjects took part.

The well-trained laboratory personnel proceeded directly to the experiment proper. Subjects from the previous experiment were again shown the prototype stimuli and asked to describe them, to ensure that no change of attitude had taken place since the last testing (this was found to be the case with three subjects, who were then eliminated from the experiment). They were then given brief practice to remind them of the task.

With the new subjects, the training procedure described in the previous section was carried out in full.

Subjects then proceeded to the experiment proper, and were instructed as follows (instructions were given in French to French-speaking subjects, and in English to the Americans who took part in the experiment.)

"I am going to show you a number of phenomena. In each case you will see the path of the red object very short at first, but I shall gradually increase it. Each time I do, I want you to tell me which of these two phenomena here [indicating the prototype stimuli] it resembles more, the one on the left or the one on the right. If it is like neither one nor the other, or if you are uncertain, say 'In between.' Do you understand?"

If the subject understood, he was told that he could look at the prototype stimuli as often as he wished, but was asked to make his decision quickly in each case, so that he would be reporting his first impression rather than his considered opinion.

In the experiment itself, the path of object B was increased in steps of 5 to 35 mm. each time. After the subject had been presented with all the stimuli, the experiment was repeated with the paths decreasing from 50 to 5 mm. in steps of 5 mm. In each case, the apparatus was covered from the subject's view when not in use and while the path lengths were being changed.

Each subject thus gave two responses for each trajectory length from 5 to 35 mm. inclusive. Since there were 40 subjects, there were altogether 80 responses at each trajectory length from 5 to 35 mm.

In order to balance out intra-serial practice effects, half the subjects were presented with the phenomena in one order, and half in the reverse order. The orders employed were as follows:—

Order 1	54:22·5	6:15	45:30	9:22·5	13·5:30	30:15
Order 2	30:15	13·5:30	9:22·5	45:30	6·0:15	54:22·5

It was appreciated that this did not control all sources of practice effects, but a design in which all these factors were controlled would have required an impossibly long time, not only on the part of the experimenter, but also of each subject (*vide infra.*).

RESULTS

Table 1 shows the raw scores for the experiment. This Table should be read as two columns, the column on the left giving the results for a descending ratio, the column on the right for an ascending ratio. Under each ratio is given the number of responses of Launching (i.e. "more like the prototype stimulus 30·0 : 3·0) "L"; the number of responses of Triggering (i.e. "more like the prototype stimuls 4·5 : 30·0") "T"; and the number of "In between" responses, "I."

Examination of Table I reveals the following points.

First, in the case of a descending ratio of speeds, increasing the length of B's path appears to have no systematic effect on the tendency to report "Triggering"; there is a slight *decrease* in the tendency to report "Launching"; the number of "In between" responses shows a slight increase.

Second, in the case of an ascending ratio, increasing the length of the path is associated with a systematic increase in the number of reports of "Triggering" and a systematic decrease in the number of reports of "Launching." The number of "In between" responses fluctuates.

TABLE I

SHOWING THE CHANGE IN DISTRIBUTION OF RESPONSES OF LAUNCHING (L), TRIGGERING (T), AND "IN BETWEEN" (I), WITH CHANGE IN LENGTH OF B'S PATH

Descending ratios					*Ascending ratios*				
Stimulus 54:22·5 (*Ratio* 2·4:1)					*Stimulus* 6:15 (*Ratio* 1:2·5)				
Length of path	*Responses*				*Length of path*	*Responses*			
	L	*T*	*I*	*Total*		*L*	*T*	*I*	*Total*
5 mm.	51	23	6	80	5 mm.	37	35	8	80
10 mm.	54	19	7	80	10 mm.	31	34	15	80
15 mm.	51	19	10	80	15 mm.	23	42	15	80
20 mm.	50	16	14	80	20 mm.	21	47	12	80
25 mm.	45	17	18	80	25 mm	22	51	7	80
30 mm.	44	17	19	80	30 mm.	15	55	10	80
35 mm	43	22	15	80	35 mm.	13	56	11	80
Total	338	133	89	560	Total	162	320	78	560
Stimulus 45:30 (*Ratio* 1·5:1)					*Stimulus* 9:22·5 (*Ratio* 1:2·5)				
Length of path	*Responses*				*Length of path*	*Responses*			
	L	*T*	*I*	*Total*		*L*	*T*	*I*	*Total*
5 mm.	49	19	12	80	5 mm.	52	21	7	80
10 mm.	50	17	13	80	10 mm.	47	21	12	80
15 mm.	47	22	11	80	15 mm.	38	30	12	80
20 mm.	45	23	12	80	20 mm.	29	35	16	80
25 mm.	46	22	12	80	25 mm.	23	46	11	80
30 mm.	40	23	17	80	30 mm.	20	51	9	80
35 mm.	45	24	11	80	35 mm.	18	49	13	80
Total	322	140	88	560	Total	227	253	80	560
Stimulus 30:15 (*Ratio* 2:1)					*Stimulus* 13:5:30 (*Ratio* 1:2·25)				
Length of path	*Responses*				*Length of path*	*Responses*			
	L	*T*	*I*	*Total*		*L*	*T*	*I*	*Total*
5 mm.	60	12	8	80	5 mm.	48	18	14	80
10 mm.	63	10	7	80	10 mm.	35	29	16	80
15 mm.	64	9	7	80	15 mm.	30	32	18	80
20 mm.	61	10	9	80	20 mm.	16	45	19	80
25 mm.	59	15	6	80	25 mm.	12	52	16	80
30 mm.	56	11	13	80	30 mm.	7	61	12	80
35 mm.	51	14	15	80	35 mm.	11	56	13	80
Total	414	81	65	560	Total	159	293	108	560

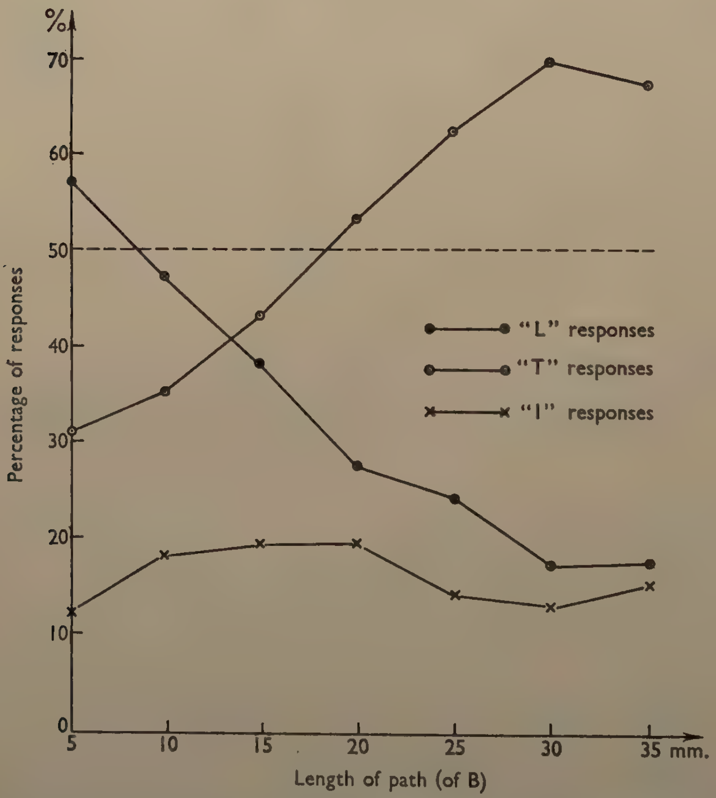
The results for the two sets of ratios, combined and converted to percentages, are given in Table II. Since there were three ratios ascending and three descending, and since as pointed out there were 80 responses at each trajectory length for each stimulus, these percentages are worked out on a base of 240. The results for ascending ratios are shown graphically in Figure 1.

TABLE II

DISTRIBUTION OF RESPONSES FOR EACH SET OF RATIOS, COMBINED AND CONVERTED TO PERCENTAGES

Descending ratios				*Ascending ratios*			
Length of path	*Percentages*			*Length of path*	*Percentages*		
	L	*T*	*I*		*L*	*T*	*I*
5 mm	66·5	22·5	11·0	5 mm.	57·0	31·0	12·0
10 mm.	69·5	19·0	11·5	10 mm.	47·0	35·0	18·0
15 mm.	67·5	21·0	11·5	15 mm.	38·0	43·0	19·0
20 mm.	65·0	20·0	14·5	20 mm.	27·5	53·0	19·5
25 mm.	62·5	22·5	15·0	25 mm.	24·0	62·0	14·0
30 mm.	58·5	21·0	20·5	30 mm.	17·5	69·5	13·0
35 mm.	58·0	25·0	17·0	35 mm.	17·5	67·0	15·5

FIGURE 1



Composite Results for Ascending Ratios, Expressed in Percentages (See Table II).

Discussion

The full implications of the present experiment cannot be appreciated without complete familiarity with Michotte's work, a discussion of which would be out of place in a paper of this nature. However, a number of observations can be made.

First, the results appear to confirm the experimental hypothesis. In the case of a descending ratio, increasing the length of B's path did not increase the tendency to report "Triggering." There was, as expected, a decrease in the number of reports of "Launching." The increase in the number of "In between" responses may be taken as evidence of the confusion brought about by the conflict of forces of organisation working in opposite directions.

In the case of an ascending ratio, the forces of organisation are working in the same direction, with the results that increase in the length of B's path is associated with an increase in the number of responses of "Triggering." In this case, the decrease in the number of "Launching" responses at greater lengths of B's path is much more marked. The combined results (Table II) show that the percentage of "In between" responses reaches a peak roughly in the same neighbourhood as the point where the curves for "Triggering" and "Launching" intersect. (See also Fig. 1.)

The relationships of the curves resemble those in the familiar judgement situation, where the response categories are, for example, "heavier than," "lighter than" and "equal to." It was in an attempt to establish this type of situation that subjects were presented with prototype stimuli and asked to compare the experimental stimuli with them.

However, there are certain dissimilarities between the conventional situation and our own that must be stressed. In the case of lifted weights, the responses are understood in the same way by all subjects, even if their actual judgements of which weights are heavier than, lighter than, or equal to the standard, differ. In the present case, among subjects who agreed in saying that a given experimental variable was "like" a prototype stimulus, the particular similarities found may have differed widely from one subject to another. To discover and define these differences one would have to question subjects extensively. We do not deny the value of the interview technique in experimental psychology, but in the present case this was irrelevant to our purpose, which was to demonstrate a relationship between changing stimulus conditions and a change in distribution of a limited number of responses.

It will be noted that, even with ascending ratios, the "Launching" response did not disappear entirely, even at the extreme length of path. It is not clear, however, whether the final value of about 17 per cent. represents an asymptotic value. In Michotte's opinion this was so, but it can be shown that the first six points fit a straight line. The seventh point, in the cases of "Triggering" and "In between" responses also, does not appear to fit the same curve as the earlier points of the graph.

A possibility here is that, after 30 mm., some other factor enters into the situation. This factor will, very probably, be the radius of action which as was mentioned in the introduction will limit the length of path in an experiment of this sort. 30 mm. appears to be the upper limit in this case. It is possible that the "Launching" responses will disappear entirely, if the length of path is increased beyond this point, but more likely that a new set of forces of organisation will come into operation. However, more experimentation is necessary before a decision can be made on this point.

It was mentioned in the Section on Procedure that about 50 per cent. of the subjects had to be rejected on the basis of the preliminary screening test, either because their replies were completely ambiguous or because they described the prototype Launching disc in terms appropriate to Triggering, and the prototype Triggering disc in terms appropriate to Launching. The reversal of the expected responses was commoner than the ambiguity, and appears to conflict with Michotte's reports of the near universality of agreement of response. However, it must be pointed out that only two ratios were used in the prototype stimuli, and that these may not have been the most apropriate for establishing the impressions. The case of A : B = 30·0 : 3·0 in particular is likely to be disturbing, as the most suitable ratio for the impression of Launching is about 3 or 4 : 1. Consequently, the particular prototype stimuli used may have tended to cause ambiguity of impression, and the response would have depended on the subject's perceptual attitude. This is supported by the case of the three subjects in the Main Experiment (mentioned in the section on Procedure), who responded differently on different occasions. However, this question could be the subject of an inquiry in its own right, and we mention it only to warn against the danger of drawing conclusions from this particular observation.

The present experiment differed in an important way from Michotte's experiments. As has been pointed out, Michotte asked subjects to describe their "impressions," and put no further restrictions on the response categories. In the major part of the present investigation, subjects were asked to compare the experimental stimuli with the prototype stimuli. Critics may object to this modification on the grounds that too much restriction is in this way imposed upon the data. This objection would be part of a much wider controversy in experimental psychology, into which it is impossible to enter here. Suffice it to say that one has the choice of extremely broad response categories, which permit one to uncover as much as possible of the subject's attitude; or narrow response categories that tell the experimenter little in themselves but which may be interpreted with the minimum of ambiguity. In order to achieve the "richness" of data that the former method permits, one has to increase the number of response categories. It is the present experimenter's opinion that the former method is more suitable for preliminary research, the latter for further experimentation which serves to make the research findings more precise.

In spite of the attempt to make the responses more precise, by limiting the scope and nature of the response categories in the present experiment, there was still room for ambiguity. Later experimenters may reduce this ambiguity yet further. Further experimentation should also be concerned to control a number of variables not controlled in the present case. *First*, the effects of order of presentation may have an important effect on the results. The ideal would have been to present the different path lengths in random order but, to do this in such a way that every path length was presented in the same relation to every other path length would have taken more time than was available to the experimenter, and more time than any subject would have been prepared to give. A re-design of the experiment, using the technique of analysis of variance, may help to solve the problem of prohibitive time. *Second*, although the results of well-trained subjects did not differ significantly from those of subjects less well-trained in the present case, this does not mean that practice has no effect upon the judgements. Training of new subjects was carried out in the present case in order to teach subjects the experimental response vocabulary; however, subjects who had been accustomed to this type of experimentation for years would certainly have approached the task with a different attitude from those experiencing it for the first time. In particular, they would already have their concepts of "Launching" and "Triggering" firmly established, and would respond without direct reference to the prototype stimuli. The present experiment was not sufficiently sensitive to detect any differences in the responses, but it is quite possible that experiments designed to take account of the effect of length of experience would be able to measure this effect, at least in a rudimentary form.

There is another, more specific criticism that may be raised against the present design. In the experiment, subjects were permitted a response of "In between" when they could not be sure which of the prototype stimuli the experimental stimulus resembled more. The controversy about the permissibility of an "uncertain" category, and the advisability of employing a forced-choice choice technique is of long standing in the history of experimental psychology, and we cannot, in this paper, justify our choice of technique at any length. Briefly, we permitted the third category because, in Michotte's opinion, if subjects did not regard the experimental stimulus as unambiguously resembling either of the prototype stimuli, the subjective impression may have been one of many not yet precisely defined. The third category, therefore, was a "blanket" category, to be broken down by future inquiry.

One final point on which future experiment seems to be required is the question of individual differences. Besides the differences, already referred to, with regard to the possible different effect of different ratios on the impressions, the whole question of exactly what the stimulus resembles, and how the resemblance may be changed by a change of perceptual attitude, requires extensive further research.

REFERENCES

MICHOTTE A. (1954). *La Perception de la Causalité.* Ed. II. Louvain.
VERNON, M.D. (1952). *A Further Study of Visual Perception.* London.
YELA, M. (1954). La nature du "Rayon d'Action" dans l'impression de causalité méchanique. *J. Psychol. norm. path.* **47-51,** 330–48.

SHORTER ARTICLES AND NOTES

GROUPING STRATEGIES WITH SIMULTANEOUS STIMULI

BY

J. A. GRAY and A. A. I. WEDDERBURN

From the Institute of Experimental Psychology, University of Oxford

Broadbent (1956) reports that two lists of digits, each presented to one ear separately so that the items in the two series coincide in time, are grouped together according to the ear-of-arrival, and that these two lists are accordingly recalled separately, one after the other. To ascertain whether such a tendency reflected some built-in mechanism or whether it was due to an optional tactic, adopted through success in making sense of message sequences in other situations, an experiment was designed in which a meaningful message would emerge for the subject if the ear-of-arrival cue was *ignored*. In this experiment, words broken up into syllables, and phrases broken up into their monosyllabic constituent words were presented to the subject, with the constituents alternating between the two ears. At the same time lists of digits were presented to whichever ear was unoccupied. The results show that recall by meaning rather than by ear-of-arrival, when these are in conflict, can occur and is no less efficient.

INTRODUCTION

Broadbent (1956 and 1958) has reported that when two lists of three digits are presented simultaneously to separate ears at a speed of two digits a second on each ear, "in the vast majority of cases . . . all the information from one channel appeared in response before any of the information from the other channel" (p. 212, 1958). This finding is peculiarly interesting because grouping by temporal order, which is the normal phenomenon at slow speeds, seemed to be unobtainable. However his material, single digits, offered no further cues for grouping other than the ear on which each item arrived. It is not clear, therefore, whether the preference shown by subjects for grouping in this way reflects some built-in mechanism, or is simply an optimal strategy for the situation—i.e. a bias on the basis of past success in classifying the input of one ear as a continuous message, when other coincident cues (e.g. message sense, voice, pitch, etc.) also fit into this classification.

If grouping-by-ear is an acquired bias, the presence of other cues *in conflict with* the ear of arrival should be able to disrupt it, and some other strategy of recall will be used.

In the following experiments, such cues were provided by (1) 3-syllable words broken up into syllables and presented to alternate ears, and (2) 3-word phrases used in the same way.

EXPERIMENT I

Equipment

A Brenell Mark IV 2-channel tape-recorder, with each channel feeding into a separate headphone, was used throughout. Subjects recorded their own responses on cards at the end of each list.

Subjects

Ten male and 5 female university undergraduates, naïve in dichotic listening situations, were given preliminary practice with one series of digits only, under the same overall conditions as in the subsequent experiment.

Material

Ten lists of digits and syllables, arranged as in the following examples, orders (1) and (2) alternating throughout:—

(1)	*Left*	*Right*	(2)	*Left*	*Right*
	EX-	6		1	CYC-
	2	TIR-		LO-	7
	PATE	9		3	STYLE

The other words used were: fol-i-o, le-pre-chaun, ob-jec-tive, mas-ter-ful, se-man-tic, cad-mi-um, ri-vu-let, hy-dro-gen. The digits were taken from a table of random digits, though no list contained any digit twice.

Procedure

The interval between the onset of each simultaneous pair was half-sec., and the lists were 13 sec. apart, which gave ample time for subjects to record their results.

Subjects were divided into two groups. Group 1 (7 subjects) was given neutral instructions, to expect syllables and digits, and to write down exactly what was heard at the end of each list. They were not told that the syllables formed words. Subjects in Group 2 (8 subjects) were told in addition that the syllables formed three-syllable English words, and were instructed to listen for these words as a unit, as well as the three digits. Both Groups were asked to write down what they recalled immediately at the end of each list, and not to "reshuffle" the items on paper, nor, as far as possible, in their heads. The second Group was also asked not to attempt to infer or reconstruct a word if they had failed to hear it as a word. Subjects were watched throughout to ensure that they did not re-order the items as they wrote them.

RESULTS

The results are given in Table I. The low mean scores indicate that all subjects found the task difficult, except Group 1 with the digits, and overall comparison of the grouping used was impossible owing to the number of ambiguous responses containing gaps and reduplicated items.

TABLE I

(*a*) PER CENT. ITEMS RECALLED

	Digits	*Syllables*	*Mean*
Group 1	78	49	64
Group 2	55	36	46

(*b*) GROUPING OF ALL-CORRECT LISTS

	Ear	*Meaning*	*Indeterminate*	*Total*
Group 1 ..	4(1.S)	—	1	5/70
Group 2 ..	1	8 (4.S)	—	9/80

But despite their higher score on individual syllables, Group 1 never made a whole word, and only had 5 lists with all 6 items correct, 4 being grouped by ear, and only the fifth one being free of phonetic approximations. Of the 9 all-correct lists of Group 2, 7 were perfectly recalled, and one had the syllables collected with a slight distortion, (CAT-ME-UM for CAD-MI-UM) without the word being recognized. Here plainly transition probabilities between syllables were not being utilized, but the differentiation between syllables and digits was alone sufficiently strong as a cue to break down the tendency to group-by-ear.

Two of the words ("leprechaun" and "rivulet") were never reported at all, perhaps owing to unfamiliarity.

We can conclude from this experiment that Group 2, owing to the greater difficulty of the task set to them, showed lower overall efficiency (fewer *items* recalled), but were more completely successful (more all-correct *lists*) when they achieved the technique of grouping-by-meaning.

Experiment 2

The second experiment was done by all subjects, in the same two Groups, immediately after Experiment 1.

Material

Ten lists of words and digits, arranged as in the following examples, orders (1) and (2) alternating throughout:—

(1)	*Left*	*Right*	(2)	*Left*	*Right*
	Mice	3		2	who
	5	eat		goes	3
	cheese	4		9	there

The other phrases used were: What the Hell, did you see, my old flame, it's in there, there she goes, tea is laid, not my sort, dear Aunt Jane.

Procedure

The timing and instructions were, *mutatis mutandis*, the same as in Experiment 1. Group 1 were told that they would hear digits and words, Group 2 that they would hear digits and words forming 3-word phrases or sentences.

Results

Higher means (Table III*a*) permit a more complete analysis of the data. First, each list was categorized for scoring as "Ear," "Meaning," or "Indeterminate" according to strict conventions. Cases where mixed strategies occurred were allotted to "Indeterminate," as were lists reported with ambiguous gaps, guesses, or reduplications. (One subject specialized in mixed strategies, e.g. "Tea 3 1 is laid 7," achieving 3 all-correct lists and the second highest total overall score in Group 1. It is noteworthy that this is exactly the strategy of temporal sequence that Broadbent's subjects found impossible.)

TABLE II

Efficiency of Performance in Relation to Method of Grouping

Method of grouping	*Number of subjects*	*Per cent total lists so grouped*	*Per cent. correct items*		
			Digits	*Words*	*Mean*
Group 1 (7 Subjects)					
Ear	4	21	80	56	68
Meaning	6	37	74	65	70
Indeterminate	—	42			
Group 2 (8 Subjects)					
Ear	5	18	71	60	66
Meaning	8	56	81	93	87
Indeterminate	—	26			

Column 2 shows the preferred method of subjects; some subjects used both methods. Column 3 shows the proportion of the total *lists* in each category. The last three columns give the percentage of individual *items* correctly recalled in each method; this provides a measure of the relative efficiency of the two methods.

Table II gives the percentage lists clearly grouped-by-ear and grouped-by-meaning, and the number of subjects using each of these methods. The *percentage success* of grouping-by-ear versus grouping-by-meaning was computed by calculating the percentage of individual items correctly reported out of the total possible within these categories; this is given for digits and words separately. It can be seen from this table that both Groups preferred grouping-by-meaning, though most subjects used both methods, and that both Groups (especially Group 2) were more successful when using this preferred method. This suggests that subjects were acting intelligently in the situation.

TABLE III

(*a*) PER CENT. ITEMS RECALLED (INCLUDING "INTERMEDIATE")

	Digits	*Words*	*Mean*
Group 1	72	53	63
Group 2	72	74	73

(*b*) GROUPING OF ALL-CORRECT LISTS

	Ear	*Meaning*	*Indeterminate*	*Total*
Group 1 ..	5 (1.S)	1	4 (1.S)	10/70
Group 2 ..	1	21 (8.S)	—	22/80

Finally in Table III, the two Groups are compared for overall efficiency (including here the "Indeterminate" category), and for the number of all-correct lists, and their method of grouping. The overall efficiency for Group 1 is not significantly different from their score in Experiment 1; for Group 2 it is much higher on both digits and words. Every subject in Group 2 achieved at least one all-correct list in the second experiment.

DISCUSSION OF RESULTS

It is clear from these results that the tendency to group simultaneously presented dichotic signals according to the ear to which they are presented can be overcome, given sufficiently strong cues favouring some other mode of grouping. This is apparent in Experiment 1, and clearly established in Experiment 2, even in the Group without instructions to follow meaning.

This system of grouping is if anything more efficient in this situation (see Tables I(*b*), II, and III(*b*)). If then we assume that the situation should require "switching attention," two possible alternatives have been suggested by Broadbent. (1) "It is possible to take a more passive attitude, and wait till after the stimuli have all arrived before beginning to deal with them successively," but "such an attitude is inefficient" (Broadbent 1957). (2) When some of his subjects did achieve immediate alternation of channels to a limited extent, he suggests "that it was done by a preliminary response to the information in some other order followed by a transposition of this order into the required one before the digits were reproduced publicly. Such a transposition is known from intelligence testing to be possible though more difficult than normal memory span" (Broadbent, 1958, p. 212). If we equate "more difficult" with "less efficient," this alternative is also ruled out here. So either our subjects were "switching," or the suggested alternatives are wrongly judged to be less efficient; or the "switching of attention" is not necessarily involved at all.

A further possibility is that subjects were utilizing transition probabilities between syllables and between words to guess at the third member of the word or phrase. We can distinguish between two ways in which this could happen. If subjects were guessing at the syllable or word they heard *last*, they must already have heard the first two syllables/words, as meaningful units, although this involved signals arriving at opposite ears. In this case, then, grouping-by-ear has already broken down before guessing can take place.

If, secondly, they were guessing at the middle syllable/word, this need not involve attention to both ears at all, as the first and last meaningful units always arrived on the same ear. However this would not account for those cases where subjects successfully recalled the *three digits* as well in the required order, for there was no way of inferring from one to another of these.

Moreover we should expect guessing to show itself in *semantic* substitutes for the correct word/syllable. Though this did occur five times in all, there were more substitutions of a word/syllable *phonetically* similar to the correct signal (nine cases), e.g. "not my salt" for "sort"—showing that the correct signal had been heard, even if imperfectly.

A final point requiring explanation is the greater difficulty for Group 2 of the task set in Experiment 1 as compared with Experiment 2. The difference is probably due to the fact that words are heard from the first as meaningful units, whereas syllables only become meaningful when the whole word is known, and are therefore less useful as immediate cues-for-grouping.

Conclusions

In terms of our original question, these results suggest that the ear-of-arrival is only one possible cue-for-grouping, although it is often dominant, especially in otherwise undifferentiated messages, as was the case for Broadbent's subjects. But by biassing the material so that meaning-cues must be followed for maximum pay-off, this skill can be displaced; the odd idiosyncratic variations, not reflected in the quantitative results, support our view that subjects are searching for an optimal strategy, and if helped by instructions, readily adopt it. Subjects are simply using whatever cues are available to interpret sensory events.

We acknowledge our thanks to Dr. J. A. Deutsch for suggesting the experiment and for supervision and many helpful criticisms, to Neville Moray for technical advice, encouragement and discussion, and to Ann Taylor for the patient and generous loan of valuable equipment and space. We also thank Professor Oldfield for providing space and facilities for the experiment.

References

Broadbent, D. E. (1956). Successive responses to simultaneous stimuli. *Quart. J. exp. Psychol.*, **8**, 145–52.

Broadbent, D. E. (1957). Immediate memory and simultaneous stimuli. *Quart. J. exp. Psychol.*, **9**, 1–11.

Broadbent, D. E. (1958). *Perception and Communication.* London.

COMBINATION OF DRIVE AND INCENTIVE*

BY

IRENE MACKINTOSH HULICKA

From the Veterans Administration Hospital, Buffalo, N.Y.

This experiment was designed to investigate the combination of drive and incentive as determinants of performance. Nine groups of rats were trained to press a bar under three levels of food deprivation (12, 24 and 36 hr.) and three incentive conditions (1, 2 and 3 pellets). Response strength was estimated by counting the number of responses with a latency of 1 sec. or less during five 20 trial sessions. The results indicated that response strength increased with hours of deprivation and with amount of food reward. Significant interactions between sessions and drive, and sessions and incentive provided additional support for the multiplicative combination of H (habit) and D (drive) and H and K (incentive). The lack of significant interaction between D and K was interpreted as supporting the hypothesis that D and K combine additively rather than multiplicatively.

This experiment was designed to investigate the manner in which the theoretical motivational variables drive (D) and incentive (K) combine with habit (H) to determine response strength (E). Evidence for the multiplicative interaction of H and D is provided by several investigations (Passey, 1948; Perin, 1942; Ramond, 1954; Williams, 1938) in which the number of reinforced trials and hours of deprivation or intensity of the unconditioned stimulus were varied. Response measures, including frequency, latency and number of responses to extinction, plotted against the number of trials show with one exception (Campbell and Kraeling, 1953) diverging curves for the different drive groups. Consequently the bulk of the evidence indicates that $E = f(H \times D)$. When D is held constant but amount or delay of reward is varied and a response measure is plotted against number of trials diverging curves are also obtained (Crespi, 1942; Dufort and Kimble, 1956; Fletcher, 1940; Grindley, 1929; Heyman, 1957; Perin, 1943) indicating that $E = f(H \times K)$. These conclusions require the assumption made by Hull (1951, 1952) and others that H is a function of the number of reinforced trials but is not affected by variations in D or K. Although there is some evidence against this assumption (Eisman, Asimow and Maltzman, 1956), more evidence is in support of it (Brown, 1956, Deese and Carpenter, 1951; Heyman, 1957; Hillman, Hunter and Kimble, 1953; Kendler, 1945; Strassburger, 1950; Teel, 1952). The limited generalization that $E = fH(D, K)$ seems permissible. Lacking experimental evidence concerning the combination of D and K, Hull (1943, 1951, 1952) speculated that they combined in a multiplicative manner. Reynolds, Marx and Henderson (1952) using resistance to extinction as a measure of response strength obtained evidence in favour of Hull's hypothesis. Recently, Spence, (1954, 1956) suggested that D and K might combine additively. An experiment by Loess (1952) may be interpreted as favouring this suggestion. Sufficient data are not available to evaluate the alternative hypotheses.

This experiment was designed to evaluate the additive and multiplicative hypotheses for the combination of D and K. Each of drive (hours of deprivation) and incentive (pellets of food) were varied three ways in a simple learning situation. If the effects of D and K on performance are unrelated, the additive hypothesis would be supported. If there is an interaction between the effects of D and K the multiplicative hypothesis would be supported.

* This research was supported by a grant from the National Science Foundation.

METHOD

Subjects and apparatus. The subjects, 90 rats purchased from the Holtsman Company, were approximately 100 days old at the beginning of the experiment. For 8 days prior to the experiment, all animals were kept on a reduced diet of 8 gm. of Purina chow per day. The apparatus, a modified Skinner box, consisted of two compartments, an entrance box and a response box. The response box was equipped with one bar, which automatically withdrew after a bar press response. An electric timer was activated when the experimenter opened the door connecting the two boxes and stopped when the animal pressed the bar.

Procedure. Prior to the introduction of the independent variables, D and K, all animals were given 50 reinforced bar press trials. All were deprived of food for 24 hr. before each of these experimental sessions. The reward for each bar press response was two 45 mgm. pellets. On each trial the animal was placed in the entrance box, entered the response box, and remained there for 40 sec. after the bar press response, sufficient time for the consumption of the pellets. The interval between trials was approximately 15 min. Four trials were given on the first day and 12 trials per day were given for the next 4 days.

TABLE I

EXPERIMENTAL DESIGN

Hours of deprivation	*Number of Pellets*		
	1 (K1)	2 (K2)	3 (K3)
12 (D12)	Group 12.1	Group 12.2	Group 12.3
24 (D24)	Group 24.1	Group 24.2	Gtoup 24.3
36 (D36)	Group 36.1	Group 36.2	Group 36.3

TABLE II

MEANS AND S.D.'S FOR FREQUENCY OF RESPONSES WITH LATENCY OF 1 SEC. OR LESS DURING 100 TRIALS

Group		*K1*	*K2*	*K3*	*Total for D Groups*
D12	Mean	30·4	38·4	42·8	37·2
	S.D.	5·5	5·2	7·9	8·2
D24	Mean	37·7	41·6	46·5	41·6
	S.D.	5·5	6·7	5·7	8·8
D36	Mean	41·6	46·4	48·3	45·4
	S.D.	7·7	6·8	4·5	7·3
Total for K groups	Mean	36·6	42·1	45·9	41·5
	S.D.	7·8	7·3	6·4	8·3

After the five days of preliminary training the 90 rats were assigned to 9 groups of 10 each. The groups were equated on the basis of response latency during the final 24 trials of preliminary training. Training was continued with variations in hours of deprivation (D) and pellets (K) as indicated in Table I. During this phase of the experiment appropriate reductions were made in the amount of food given in the home cage to account for the number of pellets given in the experimental situation. Excluding these reductions, all groups were fed 8 gm. immediately after each experimental session. Twelve, 24 or 36 hr. preceding the next experimental session, another 8 gm. were supplied. After the introduction of D and K variations, each animal was given 20 spaced trials every other day for 5 days. The latency of each bar press response was recorded.

RESULTS

The results were analyzed in terms of the frequency of trials for which the response latency was 1 sec. or less. Table II presents the means and S.D. for this measure of response strength for the 100 trials after the introduction of D and K variations.

TABLE III

ANALYSIS OF VARIANCE OF FREQUENCY OF RESPONSES WITH LATENCY OF 1 SEC. OR LESS

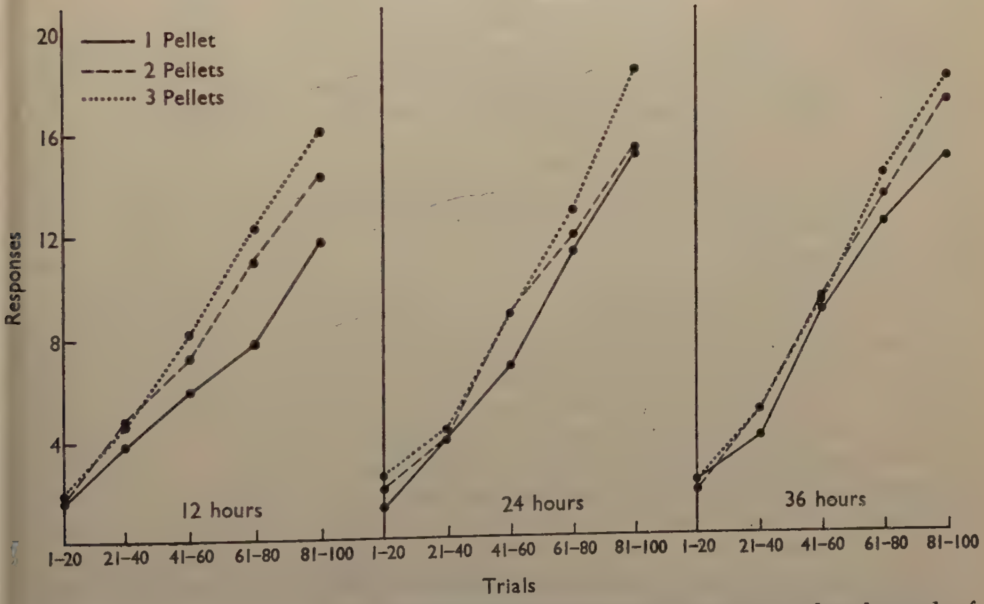
Source	*d.f.*	*Mean Square*	*F*	*P*
Between drives	2	512	11·58	<0·01
Between incentives	2	657	14·86	<0·01
Incentive × drive	4	26	0·59	>0·05
Within	81	44·2	—	—

TABLE IV

ANALYSIS OF VARIANCE OF FREQUENCY PER 20 TRIALS OF RESPONSES WITH LATENCY OF 1 SEC. OR LESS

Source	*d.f.*	*Mean Square*	*F*	*P*
Between drives	2	1024	30·1	<0·01
Between incentives	2	1314	38·5	<0·01
Between sessions (20 trials) ..	4	27982	824·0	<0·01
Drive × incentive	4	52	1·5	>0·05
Drive × sessions	8	160	4·5	<0·01
Incentive × sessions	8	144	4·2	<0·01
Drive × incentive × session	16	34	—	—

FIGURE 1



The mean frequency per 20 trials of responses with a latency of 1 sec. or less for each of 9 experimental groups.

Table III shows the results of analysis of variance when applied to these data. Figure 1 summarizes for each of the 9 groups the mean frequency per 20 trials of responses with latency of 1 sec. or less. In Figure 1 the three separate graphs represent the three drive conditions. A similar graph could be prepared to represent the three incentive conditions by rearranging the curves. The results of analysis of variance applied to the data in Figure 1 are given in Table IV. A separate analysis was done to determine whether the differences between drive groups and incentive groups during the first 20 trials were significant. The obtained F's did not approach statistical significance.

Discussion

The results indicated, as would be expected on the basis of other investigations, that response strength as measured by the frequency of responses with a latency of 1 sec. or less, was a function of trials, hours of deprivation and amount of reward. Since the bar press response was not well established prior to the introduction of the D and K variations, the differences between experimental sessions reflects the growth of H. The numerous investigations which indicate that H is not a function of D or K make it permissible to assume that H increased equally for all groups. Therefore, the significant interactions between incentives and sessions and between drives and sessions, and the diverging curves shown in Figure 1 supply additional evidence in favour of the multiplicative combination of H and K and of H and D.

The purpose of this experiment was to provide information concerning the combination of D and K. If D and K combine multiplicatively, the differences between K groups should vary with D and vice versa. If D and K combine additively, differences between K groups should not vary with D and vice versa.

For example,

$$\begin{array}{ll} \text{If } E = fH(D \times K) & \text{If } E = fH(D + K) \\ E\,12.2 - E\,12.1 = HD\,12\,(K2 - K1) & E\,12.2 - E\,12.1 = H(K2 - K1) \\ E\,24.2 - E\,24.1 = HD\,24\,(K2 - K1) & E\,24.2 - E\,24.1 = H\,(K2 - K1) \end{array}$$

Thus if $E = fH(D \times K)$ the interaction between D and K should be significant. The fact that the obtained interaction is not significant may be interpreted as evidence against the multiplicative and for the additive hypothesis for the combination of D and K.

These results are somewhat at variance with those reported by Reynolds, *et al.* (1952). They found that during training their high drive—high reward group (our Group 36.3) exhibited longer response latencies than the low drive—high reward group (our Group 12.3). Our results are not in agreement with this portion of their findings. They, however, reported that their low drive—low reward (our Group 12.1) exhibited longer response latencies than the high drive—low reward group (our Group 36.1). Here, our results correspond to theirs. One reason for the different results may be that their high drive group was deprived of food for 48 hr. while ours was deprived for 36 hr. It is possible, as Spence (1956) suggests, that at the extremely high drive level, competing responses are more apt to occur.

During preliminary training all animals were given 2 pellets per bar press. For the remainder of the experiment, the animals were rewarded with 1, 2, or 3 pellets respectively. No significant difference was found between incentive groups during the first 20 trials after the change in reward. Thus, this experiment provides no evidence for abrupt changes in performance following reward switching. This finding is in agreement with that of Pereboom (1957) and provides additional evidence for his suggestion that Hull's Theorem 30 requires revision.

Although the results of this experiment may be interpreted as indicating that D and K combine in an additive manner, the limited generalizability of this conclusion should be recognized. Before definitive general statements can be made about the combination of D and K variables, a variety and range of drives and incentives should be used, with different responses, more complex learning situations, and different levels of H.

REFERENCES

BROWN, J. (1956). The effect of drive on learning with secondary reinforcement. *J. comp. physiol. Psychol.*, **49**, 254–60.

CAMPBELL, B. A., and KRAELING, D. (1953). Response strength as a function of drive level and amount of drive reduction. *J. exp. Psychol.*, **45**, 97–101.

CRESPI, L. P. (1942). Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, **55**, 467–517.

DEESE, J., and CARPENTER, J. A. (1951). Drive level and reinforcement. *J. exp. Psychol.*, **42**, 236–8.

DUFORT, R. H., and KIMBLE, G. A. (1956). Changes in response strength with changes in the amount of reinforcement. *J. exp. Psychol.*, **51**, 185–91.

EISMAN, F., ASIMOW, A. and MALTZMAN, I. (1956). Habit strength as a function of drive in a brightness discrimination problem. *J. exp. Psychol.*, **52**, 58–64.

FLETCHER, F. M. (1940). Effects of quantitative variation of food incentive on the performance of physical work by chimpanzees. *Comp. Psychol. Monogr.*, **16**, No. 82.

GRINDLEY, G. C. (1929). Experiments on the influence of amount of reward on learning in young chickens. *Brit. J. Psychol.*, **20**, 173–80.

HEYMAN, W. (1957). Certain relationships between stimulus intensity and stimulus generalization. *J. exp. Psychol.*, **53**, 239–48.

HILLMAN, B., HUNTER, W., and KIMBLE, C. (1953). The effect of drive level on the maze performance of the white rat. *J. comp. physiol. Psychol.*, **46**, 87–9.

HULL, C. L. (1943). *Principles of Behavior.* New York.

HULL, C. L. (1951). *Essentials of Behavior.* New Haven.

HULL, C. L. (1952). *A Behavior System.* New Haven.

KENDLER, H. H. (1945). Drive interaction: II. Experimental analysis of the role of drive in learning theory. *J. exp. Psychol.*, **35**, 188–98.

LOESS, H. B. (1952). The effect of variations of motivational level and changes in motivational level on performance in learning. *Unpublished doctoral dissertation, State Univ. of Iowa.*

PASSEY, G. E. (1948). The influence of intensity of unconditional stimulus upon acquisition of a conditional response. *J. exp. Psychol.*, **38**, 420–8.

PEREBOOM, A. C. (1957). An analysis and revision of Hull's Theorem 30. *J. exp. Psychol.*, **53**, 234–7.

PERIN, C. T. (1942). Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, **30**, 93–113.

PERIN, C. T. (1943). A quantitative investigation of the delay of reinforcement gradient. *J. exp. Psychol.*, **32**, 37–51.

RAMOND, C. K. (1954). Performance in selective learning as a function of hunger. *J. exp. Psychol.*, **48**, 265–70.

REYNOLDS, B., MARX, M., and HENDERSON, R. (1952). Resistance to extinction as a function of drive-reward interaction. *J. comp. physiol. Psychol.*, **45**, 36–42.

SPENCE, K. W. (1954). Current interpretations of learning data and some recent developments in stimulus response theory. In *Learning Theory, Personality Theory and Clinical Research.* New York.

SPENCE, K. W. *Behavior Theory and Conditioning.* New Haven.

STRASSBURGER, R. C. (1950). Resistance to extinction of a conditioned operant as related to drive level at reinforcement. *J. exp. Psychol.*, **40**, 473–87.

TEEL, K. S. (1952). Habit strength as a function of motivation during learning. *J. comp. physiol. Psychol.*, **45**, 188–91.

WILLIAMS, S. B. (1938). Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, **23**, 506–22.

BOOK REVIEWS

Effect of Small Doses of Alcohol on a Skill Resembling Driving. By G. C. Drew, W. P. Colquhoun and Hazel A Long. Medical Research Council Memorandum No. 38. London. H.M.S.O. 1959. Pp. 108. 7s. 6d. net.

The purpose of this inquiry was to investigate the effects of small doses of alcohol upon sustained performance on the "Miles Motor Driving Trainer"—a complex tracking task presumed to have a good deal in common with driving. Forty subjects, average age 31 years and all with driving experience, took part in the experiments, which extended over five separate 2-hr. sessions. On each session the subjects received one of four alcohol doses or a placebo dose, the highest alcohol dose being roughly equivalent to 5 fluid ounces of whisky for a man of average weight. Subjects were not, however, informed as to which of the five types of dose they had received. The condition of the stomach was kept as constant as possible and blood, urine and breath samples were taken at half-hourly intervals. A series of personality tests was also administered.

The principal results were as follows: (1) Mean error increased significantly with increasing blood alcohol; with a blood alcohol concentration of some 80 mg. per cent. the deterioration in accuracy amounted to about 12 per cent. (2) The shape of the mean error curve was similar to that of the blood and urine alcohol curves. (3) The average speed of "driving" showed no significant change with increasing blood alcohol. (4) Individual differences were rather marked and personality assessments, especially those relating to introversion-extraversion, showed a definite relation to the behaviour changes.

Professor Drew and his colleagues comment sensibly on these findings and show exemplary caution in arguing from them to driving under ordinary conditions. They provide a most useful analysis of changes in the various components of performance (which may or may not lead to decrement in overall achievement) and of the incidence and nature of compensatory activities. They also lay proper stress on temperamental factors in relation to individual differences in reaction to alcohol.

It is obvious that the results of this study have close—perhaps decisive—relevance to a vexed social issue. At the same time, it is far from easy to translate the findings into real-life equivalents. In the first place, subjects were not told whether they had been given alcohol (and if so, how much) and for this reason might be expected to perform with somewhat excessive caution. This might affect particularly the results obtained under placebo conditions. In the second place, the experimental situation clearly provided little or no opportunity for the exercise of judgement or decision, which may well prove of far greater significance than formal tracking-ability in regard to the effects of alcohol upon driving skill. And in the third place, it might seem a pity that reliance in assessing individual differences was placed almost wholly upon rating scales and personality inventories, the limitations of which are widely acknowledged. Although we may welcome this intrusion of personality study into the analysis of skill, one may hope that it will develop along lines somewhat less constricted and more realistic than those advocated here.

This is a competent and level-headed report in the tradition of W. H. R. Rivers and H. M. Vernon. Sophisticated in design, the experiments measure up to the highest standards of contemporary psychology. The authors have taken full advantage of the advances in method that grew out of the war-time work of the Cambridge Psychological Laboratory and have applied to good effect the newer notions of skill and its deterioration which we owe above all to Sir Frederic Bartlett. The next step, perhaps, is to attempt to envisage these processes in concrete neurological terms.

O. L. Zangwill.

The Anatomy of Judgment. By M. L. Johnson Abercrombie. London. Hutchinson. 1960. Pp. 156. 25s. net.

This is a most interesting little book, written by a zoologist, which describes an experimental attempt to evoke freer and more fruitful ways of observing and thinking among young medical students. In a series of discussion classes, Mrs. Abercrombie started a number of likely hares (such as the meaning of the word "normal" in an anatomical text, or the facts and implications to be gleaned from a pair of radiographs of the child and adult human hand). The resulting wrangles bore some of the marks of a group analytic situation and involved feelings and attitudes not always comfortable or immediately satisfying to the participants. A mass of fascinating data interesting to the unprejudiced student of observation, thinking and judgement was thrown up, which the author scrutinizes with imagination and insight within a framework of contemporary psychological concepts. She succeeds, without labouring theoretical interpretations, in making skilful and effective use of ideas deriving from Ames, Bartlett and Foulkes, though it is disappointing that she does not press her analysis a little nearer to the essence of judgement and decision. Anybody concerned with university teaching will recognize that the fresh ground she breaks will have, before long, to be much more deeply cultivated. It offers, as Mrs. Abercrombie has shown, a grand opportunity for the concrete study of the "higher mental processes" in genesis and action. Any teacher in whom this book arouses disquiet will gain from it. Psychologists who read it will find double cause for gratitude—and perhaps a spur to go and do likewise.

R. C. Oldfield.

Handbook of Physiology. Section 1: *Neurophysiology. Volume* 1. Editor-in-Chief: John Field. Section Editor: H. W. Magoun. Executive Editor: Victor E. Hall. American Physiological Society (Sole Agents: The Williams & Wilkins Co., Baltimore). Pp. xiii + 779. 1959. $22.00.

This monumental venture on the part of the American Physiological Society aims to provide a systematic and authoritative compendium of present-day physiological knowledge. The first section is devoted to neurophysiology and is to comprise three volumes, of which this is the first. It is understood that the remaining two volumes are in active preparation and are due to appear in the near future.

The present volume consists of thirty odd chapters loosely grouped in four parts. The first part is concerned with problems of neuron physiology, nervous conduction and neuromuscular transmission and is fittingly introduced by Prof. J. C. Eccles. The papers in the second part deal with brain potentials, intrinsic rhythms, evoked responses and some aspects of epilepsy; this section is introduced by Prof. A. Fessard. Lord Adrian leads off the third section, concerned with sensory mechanisms and the special senses; his short introductory essay is a model of relevance and clarity. The final section is wholly devoted to vision and is ably introduced by Prof. H. K. Hartline. Mention should also be made of the erudite opening chapter by Dr. Mary Brazier on the historical development of neurophysiology.

It will be denied by no one that Dr. Magoun has chosen his team of contributors extremely well. In general, the authors succeed admirably in the difficult task allotted them and only in rare cases does critical judgment appear deficient. If fault there be, it is perhaps shown in a certain tendency to overstress work too recent for it to be placed in proper perspective. This is correlated with the omission of much earlier work still accepted as standard. Although neurophysiology is advancing so fast that a correct perspective is often difficult, a more conservative attitude might well have resulted in a better balanced text. After all, a Handbook is—or should be—an authoritative definition of the status of its subject and the secure repository of the work of a generation. Much of the present volume qualifies in this sense, though not perhaps quite all of it.

Experimental psychologists will find this text invaluable. The chapters on the senses, in particular, provide an up-to-date account of modern work of the greatest interest and value. At the same time, it should be said that—with some notable exceptions—there is little direct reference to the findings of psychological experiment. It is understood, however, that contributions of more immediate relevance to psychology are to be included in the later volumes.

It is the evident hope of its sponsors that this new venture will find a place in the long and praiseworthy tradition that extends from Haller's *Elementa* to the formidable multi-volume German handbooks of the nineteenth and early twentieth centuries. There is, however, one unfortunate difference. The prevailing division of the modern world has led to practically no account being taken here of contributions to neurophysiology from the other side of the Iron Curtain. Unless the omission is significantly remedied in later volumes, this *Handbook* will inevitably come to be regarded over large sections of the globe as a Handbook of Western Physiology. To those who believe in the unity of science this might seem a pity.

This volume is a major contribution to science and scholarship. All who have been concerned in its production are to be congratulated on a remarkable achievement.

O. L. Zangwill.

Research Techniques in Human Engineering. By Alphonse Chapanis. Baltimore. Johns Hopkins Press. 1959. Pp. vii + 316. $6.

This is a book whose purpose is avowedly to present methods rather than results. As such it fills a definite need since much of the application of psychology to practical problems requires an experiment for each problem rather than the use of knowledge picked out of a reference book. In many cases, therefore, there is a need for people working in industry to acquire the research approach, and to learn to apply experimental methods to their own problems. Indeed there are too few psychologists to carry out all the experimental work which needs doing on simple *ad hoc* problems; even if this were not a misuse of an academic training in the subject. One may therefore hope that Professor Chapanis' book may help to disseminate the experimental approach to behaviour.

The various sections of the book deal with methods of direct observation (activity sampling and other work study techniques), with methods of studying accidents (such as the critical incident technique) with statistics, with experimental design, with general research strategy, with the psycho-physical methods, and with articulation testing techniques. As this list shows, the treatment is fairly exhaustive, and the presentation of each section is clear and accurate. Each general point is illustrated by a concrete example of an investigation to which it was relevant. There should be no difficulty for engineers and other non-biologists in absorbing the information Professor Chapanis provides.

The subject matter would doubtless have been rather different in a book produced on this side of the Atlantic. In England applied psychology tends to be associated with applied physiology more closely than it is with work study; so that one would find it surprising to meet a complete list of the Therblig categories in a book which contains no information on the measurement of energy consumption at all. Equally, the allocation of a complete chapter to articulation testing does not reflect the relative importance of that type of problem in British application. Some of us may also regret the subject matter of the statistical chapters, since they provide a clear exposition solely of methods based on the analysis of variance. Recent criticism of such methods has been so severe that many investigators are abandoning the use of Latin Square designs, and are using non-parametric methods wherever possible. Even if this is felt to be an extreme course, it would surely be worth giving such simple methods as the Sign Test in a book which goes so far as to consider the extraction of fourth order interactions.

But it is likely that, quite apart from cultural differences, any individual applied psychologist would produce a different list of topics to be included in a book with this general aim. Professor Chapanis is to be congratulated on his success in producing a compendium of research techniques which is simple enough for the complete novice, and yet sufficiently detailed to be of value for reference by the professional.

D. E. BROADBENT.

What is Cybernetics? By G. T. Gilbaud. (Translated by Valerie Mackay from *La Cybernetique.* Paris, Presses Universitaires de France, 1954.) London. Heinemann: Contemporary Science Series. 1959. Pp. viii + 126. 10s. 6d.

A non-technical introduction to cybernetics has long been needed, and this reasonably priced volume should go far to fill the gap.

Most of the book is devoted to a general discussion of information, its measurement, and the laws of communication, with several useful examples worked out numerically. Feedback and network theory are dealt with more shortly, and no account is given of the actual properties of closed-loop systems. In particular, no mention is made of the important effects of delayed feedback. The theory of games is very briefly mentioned in the final chapter.

The selection and treatment of topics, and the authorities cited, will seem refreshingly different to the Anglo-Saxon reader. Most of the important cybernetic ideas seem to have appeared quite early in French mathematical and philosophical writing, to be embodied later (often unwittingly) in English and American engineering. To judge from this book the French still prefer a broad intellectual synthesis rather than a concrete problem-solving approach.

The translation is excellent, though certain technical terms have been transliterated rather than translated. For instance, Figure 2 (p. 13) would normally be called a "switch" rather than an "interrupter."

E. R. F. W. CROSSMAN.

ZEITSCHRIFT FÜR EXPERIMENTELLE UND ANGEWANDTE PSYCHOLOGIE

Organ der Deutschen Gesellschaft für Psychologie

Herausgeber für Deutschland	Prof. Dr. Ph. Lersch, München Prof. Dr. F. Sander, Bonn
Herausgeber für Österreich	Prof. Dr. Dr. H. Rohracher, Wien
Herausgeber für die Schweiz	Prof. Dr. R. Meili, Bern
Redaktion	Dr. C. J. Hogrefe, Göttingen

Band VII/1

ADOLF OTTO JÄGER, Einige emotionale, conative und zeitliche Bedingungen des Erinnerns.

ADRIENNE MILER, Vergleich der Vergessenskurven für Reproduzieren und Wiedererkennen von sinnlosem Material.

KURT MÜLLER, Die Beziehungen zwischen wahrgenommenen und objektiven Geschwindigkeiten von Translationsbewegungen.

WALTER KRISTOF, Ein Umschlag der Heringschen Täuschung bei Abwandlung der Vorlage.

RUDOLF POPHAL und ERICH DUNKER, Zeitlupenstudien des Schreibvorganges.

JACQUES EX, Situationsanalyse und sozialpsychologisches Experiment.

KONRAD JOERGER, Das Erleben der Zeit und seine Veränderung durch Alkoholeinfluß.

WALTER KRISTOF, Die Propellertäuschung als Effekt stroboskopischer Beleuchtung.

FREDERIC C. BARTLETT, A reply to Dr. Kristof.

All articles are provided with summaries in English and French.

The journal appears quarterly and contains 160 pages per number.

Subscription is DM 64,—per volume.

VERLAG FÜR PSYCHOLOGIE, DR. C. J. HOGREFE, GÖTTINGEN
(Brentanoweg 10, Germany.)

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

Papers submitted for publication and all other communications on Editorial matters should be addressed to "The Editor, *Quarterly Journal of Experimental Psychology*, Psychological Laboratory, Downing Street, Cambridge, England."

Only original papers are accepted for publication. The copyright in any material accepted for publication is retained by the *Journal*. Reproduction in whole or in part is forbidden, except (1) under terms of the *Fair Copying Declaration* of the Royal Society, or (2) when specific permission has been given by the Editor.

Intending contributors should pay attention to the conventions of the *Journal*, and especially to the form in which references are given. Manuscripts should be typewritten with double spacing and with wide margins. Diagrams, which should be kept to the minimum, must be submitted in a form suitable for reproduction without redrawing. They should be drawn in black ink on heavy paper; lettering should be inserted not on the diagram but on a tracing fixed over the diagram. Legends and tables should be typed on separate sheets of paper.

Contributors receive 25 copies of their paper free. If they require further copies they should apply to the Editor at the time of returning proofs.

CONTENTS

P. C. WASON. On the Failure to Eliminate Hypotheses in a Conceptual Task - - - - - - - - - - - - 129

F. H. GEORGE and S. B. McINTOSH. Experimental Disorientation and Conceptual Confusion - - - - - - - - 141

ALAN STANILAND. Redundancy as an Experimental Variable - 149

PETER LADEFOGED and DONALD E. BROADBENT. Perception of Sequence in Auditory Events - - - - - - - 162

D. G. BOYLE. A Contribution to the Study of Phenomenal Causation 171

Shorter Articles and Notes

J. A. GRAY and A. A. I. WEDDERBURN. Grouping Strategies with Simultaneous Stimuli - - - - - - - - - 180

IRENE MACKINTOSH HULICKA. Combination of Drive and Incentive - - - - - - - - - - - - 185

Book Reviews - - - - - - - - - - - 190

The Quarterly Journal of Experimental Psychology is published for the Experimental Psychology Society by W. Heffer & Sons Ltd., Petty Cury, Cambridge, England, to whom all business communications should be addressed.

The subscription is £2 10*s*. 0*d*. (post free) per volume of 4 numbers. Numbers appear quarterly. The cost of any single number is now 16*s*. 0*d*. (post free).

Undergraduates and graduates of less than three years' standing are offered a special rate of £1 5*s*. 0*d*. per volume. They should apply to the Hon. Treasurer, Experimental Psychology Society, Department of Psychology, The University, Exeter, enclosing a remittance for £1 5*s*. 0*d*. and stating in which university department they have worked and what year they expect to graduate or did graduate.

All other subscriptions and orders should be placed with W. Heffer & Sons Ltd., or any bookseller.

PRINTED BY W. HEFFER & SONS LTD., CAMBRIDGE, ENGLAND